

REPORT OF THE
**Planning Group on Commercial Catch, Discards
and Biological Sampling**

**Lisbon, Portugal
5–8 February 2002**

This report is not to be quoted without prior consultation with the General Secretary. The document is a report of an expert group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the views of the Council.

International Council for the Exploration of the Sea

Conseil International pour l'Exploration de la Mer

Palægade 2–4 DK–1261 Copenhagen K Denmark

TABLE OF CONTENTS

Section	Page
1 INTRODUCTION.....	1
1.1 Terms of reference.....	1
1.2 List of participants.....	1
1.3 Background.....	1
2 WORK PLAN.....	2
3 REPORT FROM SUB-GROUP A.....	2
3.1 Landings.....	2
3.2 CPUE data.....	3
4 REPORT FROM SUB-GROUP B.....	3
4.1 Length and age sampling of national commercial landings.....	3
4.2 Length and age sampling of foreign commercial landings.....	4
4.3 Sampling of other biological data (SWALMF).....	5
4.4 Length and age sampling of discarded catch.....	5
5 REPORT FROM SUB-GROUP C.....	6
5.1 Data storage procedures and software for data aggregation.....	6
5.2 Data exchange formats.....	7
5.3 General considerations.....	7
5.4 Cost implications.....	8
6 INTERNATIONAL CO-ORDINATION.....	8
6.1 Sampling in 2002.....	10
7 RECOMMENDATIONS.....	10
8 ACKNOWLEDGEMENT.....	11
9 REFERENCES.....	11
APPENDIX 1.....	17
APPENDIX 2.....	18
APPENDIX 3.....	19
APPENDIX 4.....	20
APPENDIX 5.....	24
APPENDIX 6.....	57
APPENDIX 7.....	63
APPENDIX 8.....	65

1 INTRODUCTION

1.1 Terms of reference

During the Annual Science Conference (89th Statutory Meeting) in Oslo September 2001 it was decided that an ICES Planning Group on Commercial Catch, Discards and Biological Sampling [PGCCDBS] should meet in Lisbon on 5–8 February 2002 to:

- a) evaluate the commercial catch (landings), discard and biological sampling programmes being implemented in the Baltic Sea, North Sea, Western and Southern waters, and assess whether or how these require ICES co-ordination;
- b) evaluate the need for developing sampling methodology, calculation methodology, data storage procedures, and software for aggregating national catch-at-age data to international catch-at-age data in a form suitable for assessment working groups, and prepare relevant proposals and work plans.

1.2 List of participants

The meeting was attended by:

Inaki Artetxe, Spain
Kenny Coull, UK, Scotland
Jørgen Dalskov, Denmark (chair)
Henrik Degel, Denmark
Guus Eltink, Netherlands
Steve Flatman, UK, England
Ole Folmer, Denmark
Maria Hansson, Sweden
Ernesto Jardim, Portugal
Helen McCormick, Ireland
Richard Millner, UK, England
Philippe Moguelet, France
Andrew Newton, UK, Scotland
Rick Officer, Ireland
Graça Pestana, Portugal
Antonio Punzon, Spain
Jukka Pönni, Finland
Tiit Raid, Estonia
Dália Reis, Portugal
Reidar Toresen, Norway
Valentin Trujillo, Spain

1.3 Background

The ICES fisheries advice critically depends on the quality of data from the commercial fisheries. The quality of these data has not in all cases been satisfactory and ICES has raised this point repeatedly. For the last 6-8 years, the EU Commission has financially supported several projects, whose objectives have been to support the Common Fisheries Policy CFP. Several projects dealt with improvement of data collection. Other projects (EMAS, Study Contract 94/013, IBSSP I, IBSSP II, FIEFA, SAMFISH, MIQES) dealing with the evaluation of basic data such as market sampling and survey data that underlie stock assessment, are currently underway or have been finished. Also, there is an increasing need to address the issue of quality assurance of the biological advice. Linking these two items would draw out the methodologies and issues relating to the use of this type of information within the assessment process, and the impact on advice.

Within the latest years projects on methods for storing, combining, and raising national data to international stock-related data-sets have been carried out.

Most of these projects have been carried out in cooperation between different national fisheries research institutes, and through this cooperation international co-ordination of the work including data collection has been achieved. The current initiative by the EU Commission (DG FISH) in providing financial support for the collection of fisheries data is much welcomed. But, as the financial support is given to national data collection programmes the existing element of international co-ordination may be missed and may cause inappropriate or missing data collection of certain species.

This Planning Groups was established in order to provide focal points for coordinating sampling activities and strategies among ICES member countries.

2 WORK PLAN

In order to use the four days meeting in the most efficient way, it was agreed to establish 3 sub-groups with the following subjects:

Group A: Landings (tonnes).

Group B: Discards and biological sampling, including sampling and calculation methodology.

Group C: Data storage procedures and software for data aggregation.

The sub-groups shared the same terms of reference:

- a) Highlight and describe areas where co-ordination within the ICES community is needed and areas where the provisions according to the EU regulations for fisheries data collection are not sufficient to obtain basic data for stock assessment purposes.
- b) Describe and suggest initiatives that could be implemented to achieve better basic data for stock assessment purposes.
- c) Recommendations. What should be done within the next year(s).

3 REPORT FROM SUB-GROUP A

3.1 Landings

ACFM has for many years highlighted the problems concerning catch figures used in the various assessment working groups. The assessments carried out by ICES use the best landings data available to ICES. These data are not necessarily identical with the official statistics and, where appropriate, include estimates of unreported landings as well as corrections for misallocation of landings by area and species. Despite considerable effort exerted on this problem, there is no guarantee that all instances of misreporting are discovered. In the assessments, the working groups try to estimate the total catch taken, including slipped catches, discards, landings which are not officially reported, and the composition of the industrial by-catches. These discards, slipped fish, unreported landings, and industrial by-catches vary considerably between different stocks and fisheries, being negligible in some cases and constituting important parts of the total removal from other stocks.

There are numerous problems such as unreliable catch figures, insufficient data collection of biological samples, uncertainties in age readings, etc., for example *Connolly et al. ICES CM 2001/P:2*, *Maxwell et al. ICES CM 2001/P:11*, *O'Brien et al. ICES CM 2001/P:13*.

In this section the sub-group has focused on:

- Mis-identification of species
- Area mis-reporting
- Undeclared landings
- Incorrect size grading

There are numerous problems that have existed for a number of years and are compounded by different standards of enforcement in the various nations; e.g. some countries have, in recent years, increased the level of enforcement to the stage where all vessels have to phone ahead with landings data; failure to land the 'telephoned' catches can result in the withdrawal of the fishing licence. However, other nations have greater difficulties with enforcement as stated in the EU Commission Staff Working Paper on "Fisheries Control in Member States" (SEC 2001 1798, 1799, 1811-1814, 1818-1824) and this situation gives scope for the problems outlined above.

To obtain more reliable landings data it is evident that better controls of landings are required in most areas.

In addition there are concerns about whether relevant stratification for areas and time is met in the EU Commission Regulation.

Most problems cannot be resolved in the short term, but two issues can be addressed almost immediately:

- The chairs of the relevant Working Groups will be approached and asked to review Commission Regulation (EC) No. 1639/2001 to ascertain whether the levels of stratification outlined in Appendices I and XII to the Commission Regulation are sufficient for their purposes.
- Introducing the use of the template examples in Appendix 1-2, created to enable each Assessment Working Group to provide an evaluation of the quality of the data for each species/stock. The completed templates will be returned to PGCCDBS. This will highlight any particular weaknesses and provide a basis for action at the next meeting of this Planning Group.

The Planning Group therefore recommends that:

- **ICES Assessment Working Groups participate in the attempts to secure adequate data for stock assessment purposes by highlighting any particular weaknesses in the quality of the data and providing a basis for action at the next meeting of this Planning Group.**

3.2 CPUE data

In many ICES assessment working groups CPUE time-series are used for tuning. According to the new EU regulation for data collection provisions for collecting CPUE data are also included. Attention should be given to the fact that according to these provisions existing time-series can/should be maintained, but no funding support for establishing new time-series will be given.

However, a review should be made of the existing CPUE time-series, which have been used during the years 1995 to 2000 by scientific assessment working groups. This review should analyse the weight given by each CPUE time-series in the final stock assessment to the corresponding abundance or partial fishing effort indices, the possibility of prolonging the corresponding times-series on the basis of disaggregated catch and effort figures, as well as the possible necessity of referring to even more detailed data.

These reviews will be evaluated by STECF, and decision on future guidelines for EU financial support for establishing new CPUE time-series will probably be taken in the autumn 2003.

4 REPORT FROM SUB-GROUP B

The need for co-ordination was considered with reference to meeting the needs for:

- Length & Age Sampling of Commercial Landings;
- Sampling of other biological data (SWALMF, *i.e.* Sex, Weight, Age, Length, Maturity, and Fecundity);
- Length & Age Sampling of Discarded Catch.

(It should be noted that the acronym SWALMF is used in this report to describe parameters as Sex, Weight, Age, Length, Maturity, and Fecundity)

Whilst the discussion was focussed upon meeting the requirements of the EU data collection regulation (1639/2001) consideration was also given to satisfying data requirements for improved ICES stock assessments.

4.1 Length and age sampling of national commercial landings

Age determination is an essential feature in fish stock assessment to estimate the rates of mortalities and growth. In order to arrive at appropriate management advice ageing procedures must be reliable. Otolith processing methods and age reading methods might differ considerably between countries. Therefore, otolith exchanges should be carried out on a regular basis, and if serious problems exist age reading workshops should be organised to solve these problems. The European Fish Ageing Network (EFAN) provides on the internet not only guidelines on how to organise otolith exchanges and workshops, but also tools for comparing the age reading results (www.efan.no under guidelines).

Otolith exchanges cannot be started for all species at the same time; therefore, the candidate species for otolith exchanges should be identified on a regional basis.

Recommendation:

- **To identify on a regional basis the candidate stocks and species requiring improved ageing.**

The data directive is requiring some laboratories to begin ageing new species. This has created several problems:

- Where there is no established local expertise in ageing new species data quality cannot be assured;
- The new levels of ageing required are sometimes so low that they do not warrant establishing a dedicated reading program in every laboratory.

EFAN data bases number 10 and 11 provide information on otolith readers who have expertise in age reading a particular fish species (www.efan.no under databases).

Recommendation:

- **To examine the possibility of sharing / transferring otoliths between laboratories.**

It was noted that the problem of concentration of ageing expertise into few laboratories must be avoided as this makes assessments vulnerable to any biases or inaccuracies in the ageing undertaken by these laboratories, and staff with expertise at these laboratories leave their jobs.

Previously, non-EU ICES members have been involved in Study Contracts that were coordinated with EU member states. The conclusion of these study contracts has created a difficulty in obtaining funds to facilitate attendance of non-EU ICES member countries in co-ordination meetings. Sampling of landings in non-EU ICES member countries are currently not coordinated.

Recommendation:

- **To involve non-EU ICES member states in the regional planning of sampling.**

Regulation 1639/2001 allows member states to claim an exemption from sampling because of low quota and landings. Co-ordination is required to check that the level of landings by states claiming exemptions does not exceed specified thresholds. This could be ensured by a group with a ToR to coordinate sampling to ensure adequate spatial and temporal sampling coverage.

The data directive also provides for reduced levels of sampling by member states to be picked up by increased sampling by other states. It is currently assumed that this will occur, but there has been no examination of the implications on data quality and coverage created by these adjustments. This is not necessarily a matter requiring co-ordination, but potentially it could have a large impact on some assessments.

4.2 Length and age sampling of foreign commercial landings

For the purposes of the sampling regulations, a foreign vessel is defined as one flying a flag which is not the same as the country of landing. Member states on whose territory landings take place are responsible for sampling all landings, irrespective of the nationality of the vessels making the landing (EC 1639/2001 section H (c)). In practice this will involve adding the tonnage landed by national vessels to those from foreign or flag vessels and applying the sampling strata from 1639/2001 Appendix XV to the total landings. In cases where the foreign landings are only a small proportion of the total landings, and sampling is randomly applied to the landings, it possible that the foreign segment may not be sampled. Furthermore, there is often considerable difficulty in sampling from foreign landings. In many cases, landings are made directly into waiting lorries for overland transport to another country. This may result in some segments of the fishery not being sampled.

Recommendation:

- **In general, the country into which fish are landed should process data sampled from the foreign fleet to produce assessment inputs scaleable to any level of landings. These data, and the raw length distributions and age estimates, should then be provided to the flag country or country of fleet origin.**
- **In cases where this is not practical or would lead to errors in derived age-compositions, the country into which fish are landed should process data sampled from the foreign fleet to produce length distributions and ensure that ageing is undertaken. These data should then be provided to the flag country or country of fleet origin.**

Achieving this will consequently require:

- co-ordination of sampling protocols (in the latter case);
- co-ordination of the timing of processing of samples / provision of data;
- for species and stocks where no local expertise in ageing currently exists it may be sensible to undertake ageing in specialist laboratories (as indicated above).

4.3 Sampling of other biological data (SWALMF)

Items that need co-ordination:

- timing of the collection of samples by species and stock during the intervals specified in Regulation 1639/2001 (in cooperation with assessment WGs),
- timing of the collation of data by species and stock,
- identification of opportunities to append biological sampling on existing sampling programs,
- establish consistent protocols for maturity staging, fecundity estimates, etc.,
- all countries should ensure that sampling for length & weight data is done by quarter where there are seasonal growth changes.

Regulation 1639/2001 allows member states to claim exemptions from the collection requirements of SWALMF where:

- Member state quotas are a small proportion of the TAC,
- Sampling is coordinated to achieve an overall level of precision.

A mechanism is required that will ensure that necessary precision levels are reached by the international data. This will require that data are combined at some stage to allow a check on precision to be made. There are major statistical problems in determining precision at the international level on combined data. Advice on how to provide such estimates of precision is required.

It must also be checked that the quota from member states claiming exemptions does not exceed the 20% threshold specified in Regulation 1639/2001.

Recommendations:

- **That regional planning occurs in the collection of these data to ensure that thresholds for data collection are met.**
- **That the collation of SWALMF occurs at a regional level.**
- **That advice is sought on methods to calculate precision levels on combined data.**
- **That for the purposes of international combination of maturity data, the maturity stages used should be mappable to a common standard such as the IBTS maturity stages (IBTS Manual, Appendix II. ICES CM 1999/D:2).**

4.4 Length and age sampling of discarded catch

It is difficult to achieve high precision in discard estimation without extremely intensive sampling, which can be cost-prohibitive. The group considered that this problem requires that discard sampling be carefully focussed to ensure the

maximum utility of the data collected. For stocks covering a broad geographical area it would be sensible to examine the possibility of distributing / focussing the sampling of member states onto particular fleets / fisheries. The advice of the fishing industry will be crucial in deciding where to target sampling.

Recommendations:

- **The ICES Study Group on Discard and Bycatch Information (SGDBI) be requested to list the planned sampling in 2002.**
- **The available data should be reviewed on a regional basis to ensure that sampling coverage reflects the relative impact of each fishery in each area.**

Inconsistencies in methodology and stratification currently prevent the full utilisation of discard estimates into stock assessments.

Recommendations:

- **That the SGDBI be asked to provide practical advice on the compatibility between fishery stratification used in sampling discards and commercial catch statistics.**
- **That countries undertaking discard sampling adhere to the sampling strategy as suggested by SGDBI.**

It is currently difficult for countries to adequately sample discards of their own fleets. The need to sample foreign and flag fleets adds an additional burden. Regulation 1639/2001 does not specify who should sample discardings from these fleets.

Recommendation:

- **That the country of origin makes arrangements to sample the discards and associated landings.**

There is currently no system for the storage of discard data in many countries. Many countries are now embarking on new discard programs. It would be sensible to take the opportunity to standardise the management of new discard databases. This issue is dealt with in the section on data handling.

Regulation 1639/2001 specifies a triennial collection of discard data for some stocks. It must be noted that such data will not be very useful in assessment. If discard data are used for recruitment indices then an estimation of discarding levels is required annually.

5 REPORT FROM SUB-GROUP C

5.1 Data storage procedures and software for data aggregation

Section on TOR b): evaluate the need for data storage procedures and software for aggregating national catch-at-age data in a form suitable for assessment working groups.

In general terms, relevant information (national datasets, documents, etc.) must be stored in an organized, robust, and consistent way, allowing updates and queries on this information in a simple and practical manner. The group feels that the framework defined by relational database systems is the most appropriate for this task.

In the following section only the problems referring to sampling information and assessment data are addressed. The group, however, recognises that there is at present still an urgent requirement to manage information associated with WG assessments (e.g., working documents and the Quality Handbook).

The group considered some possible objectives:

A) Working Group support

To support the WG stock coordinators in the aggregation of national data sets to provide age compositions and estimates of precision for use in stock assessment. At the same time, records of the procedures used in terms of data

aggregation and decisions on allocations should also be derived. Ideally this would be produced automatically by the same system.

B) Sampling management

To provide a tool for the analysis of international sampling results in order to review coverage and sampling effort allocation on a stock basis. The tool would have to include facilities for evaluating sampling effort by those strata defined in the EU data collection regulation requirement.

In an ideal system, these two objectives might be achieved using just one database system which included international data at the sample level together with catch statistics. This point is discussed below.

The group considered that there were two quite different methods of accomplishing the above objectives:

Low-level approach:

This is based on the idea that all raw international sample data and catches by fleet or fishery unit, etc., are stored in one comprehensive database. The application would need to allow for raising, aggregation, and logging procedures, and should also produce the files required for assessment software use. One such example is the BALTCOM (Appendix 3) database, which contains much of the functionality required, but would need some further development.

Advantages – consistency, completeness, flexibility in the use of data.

Disadvantages – full-commitment, complexity, confidentiality, duplication.

High-level approach:

Based on the assumption that national data sets are available in an aggregated form (time/space/fleet/species), such as the data provided to stock coordinators. In this case, the requirement is to be able to carry out standard analyses for assessment working groups (or other international groups) on the stored data, and to be able to record the procedure used. In addition, information on the sampling coverage achieved could be used to coordinate international sampling requirements, e.g. for the EU data collection regulation.

Examples of this approach include ISLDB (Appendix 3 and Appendix 4) which is mainly for storage and retrieval of sampling information (meta-data), and VPABASE (Appendix 3 and Appendix 5) which is principally for age data manipulation, including tracking of the procedures used in combining data.

Advantages – simplicity, flexibility in the level of data allowed.

Disadvantages – limited analysis, unknown procedures at national level.

An example schematic of these two approaches, based on the use of currently available (but not fully-developed) software applications is given in Figure 1.

5.2 Data exchange formats

Pastors (Appendix 6) and Eltink (Appendix 7) suggest that a simple text format would be preferable for data exchange. The planning group fully supports this approach, and a recommendation to this effect is made below.

5.3 General considerations

Systems which include international data need clearly-appreciated management and development responsibilities.

Management issues, including ownership, security, maintenance, and provision of access should be the responsibility of an international scientific organisation like ICES. This has been partially addressed by groups such as the Coordinating Group on ICES Advice (1999), which described the role of ICES in maintaining data sets and documentation of methods.

The responsibility for development of these tools should rest with the international scientific community, as current practice. The development environment should be Web-compatible and preferably using a relational database management system. Also, the source code should be available to the international scientific community following the open-source model.

In the recent past several Working Groups have discussed the need for such tools to be developed.

5.4 Cost implications

The Low-level approach would require more development and maintenance effort (including management and security of data, etc.), and consequently higher costs and a longer timescale, than the High-level approach.

Recommendations:

- **These assume that further development of these tools will be carried out to address the problems identified above, including migration to a Web environment.**

The planning group considers that there is a need for:

Trial of the BALTCOM database for the storage of discard data from the new discard programmes initiated under the EU Regulation.

Trial of VPABASE by Stock Coordinators for data aggregation and procedure logging, and report back on its potential and any required modifications.

Trial of ISLDB by Working Groups for sampling meta-analysis such as sampling coverage and intensity.

Code standards to be established and maintained by ICES, which should avoid some of the problems associated with data exchange.

A standard exchange format based on text files (delimited or XML). Variables should be clearly described and in agreement with the standard codes.

Trial of low-level databases such as BALTCOM and DATRAS (Appendix 8) for the storage of SWALMF data.

The group therefore recommends that:

- **DIFRES be requested to review the licensing and additional technical requirements for the use of the BALTCOM database by the international community, particularly with respect to discard data storage (and in cooperation with the SGDBI), and report back to the next planning group meeting (PGCCDBS).**
- **DIFRES be requested to develop a test version of VPabase, using a range of WG data, and report progress at the next PGCCDBS meeting.**
- **IPIMAR be requested to extend ISLDB as necessary to accommodate additional national data sets as discussed, and report progress at the next PGCCDBS meeting.**
- **Subject to approval at the next PGCCDBS meeting, these applications can be released for trials by stock coordinators, Working Groups, and the next Discards meeting as appropriate.**
- **ICES should establish and maintain standard codes for data to minimize problems associated with data exchange.**
- **Data exchange formats should be based on text files (delimited or XML), and variables should be clearly described and in agreement with the standard codes.**

6 INTERNATIONAL CO-ORDINATION

Whilst regulation 1639/2001 specifies the sampling levels for length and age for each stock sampled, each country can sample the fleets that it chooses. There is no international co-ordination to ensure that sampling occurs from all fleets fishing any particular stock. Furthermore, regulation 1639/2001 requires cooperation and task sharing between member states.

Previous opportunities for co-ordination have been lost through the conclusion of projects such as SAMFISH, EFAN, IBSSPII, DELASS. Whilst other opportunities to coordinate and plan sampling should be investigated, it was considered that most assessment WGs have insufficient time to consider co-ordination.

Recommendation:

- **That meetings are convened to coordinate sampling to ensure adequate spatial and temporal sampling coverage.**

Scope of international co-ordination required:

- Coordinate sampling to ensure adequate spatial and temporal sampling coverage;
- Review methodologies for collection of data;
- Agree on otolith exchange programmes;
- Agree on mechanisms for evaluation in future years.

In the future PGCCDBS meetings should include ToRs:

- To evaluate sampling in the previous year in terms of:
 - Spatial & Temporal Coverage
 - Precision levels (including age reading)
- To establish quality assurance protocols for assessment data

PGCCDBS meetings would also provide an opportunity to organise distribution of otolith reading between countries.

Whilst members of the assessment working groups are probably best placed to identify the problems in data quality affecting their assessments PGCCDBS agreed that it would not be realistic to ask assessment working groups to take on the additional task of international co-ordination and planning of sampling. In addition there are many species and stocks for which EU Regulation 1639/2001 requires co-ordination, which are not assessed by ICES working groups.

The option of separate international co-ordination meetings convened for each region was also considered to be unworkable. Such an arrangement would introduce a number of new meetings into a schedule that is already overloaded. This arrangement could also introduce a delay before the groups could meet, as the meetings would need to be approved by the convening bodies. Furthermore, there would be a need for some co-ordination between regional international co-ordination groups, and hence a need for another co-ordination meeting.

The group considered the option of undertaking the task of international co-ordination within the PGCCDBS. It was agreed that such an arrangement would retain the flexibility to convene sub-groups of PGCCDBS to work on matters of international co-ordination both intersessionally and at the PGCCDBS. This arrangement would not add a new meeting to the existing schedule.

Recommendation:

- **That ACFM establish PGCCDBS as an annual working group, and that PGCCDBS should meet in February.**

Arrangements for the collation of data on assessed stocks

The PGCCDBS discussed how best to collate the information on deficiencies in assessment to enable an evaluation of areas where co-ordination would improve data quality. Each assessment WG has one or more species coordinators who collate international assessment data. It was considered that species coordinators could report the deficiencies in the data contributing to stock assessment to PGCCDBS without a huge increase in their workload.

Recommendation:

- **That WG species coordinators be asked to report to PGCCDBS the deficiencies in the commercial data contributing to stock assessment.**

6.1 Sampling in 2002

In accordance with Commission Regulation 1639/2001 a number of species shall be subject for biological sampling. This sampling scheme is divided by areas. Table 1 is prepared in order to identify if there should be areas and/or species which normally are assessed by ICES. Unfortunately, no information from Belgium, Faroe Islands, Germany, Island, Latvia, Poland and Russia were available as no representatives from these countries participated in the PG meeting. Though not all information was available, Table 1 shows that all species for which stock assessment is carried out, sampling is conducted.

7 RECOMMENDATIONS

Recommendation concerning landings:

- **ICES Assessment Working Groups participate in the attempts on securing adequate data for stock assessment purposes by highlighting any particular weaknesses in the quality of the data and provide a basis for action at the next meeting of this Planning Group.**

Recommendations concerning biological sampling of national landings:

- **To identify on a regional basis the candidate stocks and species requiring improved age reading.**
- **To examine the possibility of sharing / transferring otoliths between laboratories.**
- **To involve non-EU ICES member states in the regional planning of sampling.**

Recommendations concerning biological sampling of foreign landings:

- **In general, the country into which fish are landed should process data sampled from the foreign fleet to produce assessment inputs scaleable to any level of landings. These data, and the raw length distributions and age estimates, should then be provided to the flag country or country of fleet origin.**
- **In cases where this is not practical or would lead to errors in derived age-compositions, the country into which fish are landed should process data sampled from the foreign fleet to produce length distributions and ensure that ageing is undertaken. These data should then be provided to the flag country or country of fleet origin.**

Recommendations concerning other biological sampling:

- **That regional planning occurs in the collection of these data to ensure that thresholds for data collection are met.**
- **That the collation of SWALMF occurs at a regional level.**
- **That advice is sought on methods to calculate precision levels on combined data.**

Recommendations concerning biological sampling of discards:

- **That for the purposes of international combination of maturity data the maturity stages used should be mappable to a common standard such as the IBTS maturity stages (Appendix II - IBTS Manual, ICES CM 1999/D:2).**
- **The ICES Study Group on Discard and Bycatch Information (SGDBI) be requested to list the planned sampling in 2002.**
- **The available data should be reviewed on a regional basis to ensure that sampling coverage reflects the relative impact of each fishery in each area.**
- **That the SGDBI be asked to provide practical advice on the compatibility between fishery stratification used in sampling discards and commercial catch statistics.**
- **That countries undertaking discard sampling adhere to the sampling strategy as suggested by SGDBI.**
- **That the country of origin makes arrangements to sample the discards and associated landings.**

Recommendations concerning data storage and data aggregation:

- DIFRES be requested to review the licensing and additional technical requirements for the use of the BALTCOM database by the international community, particularly with respect to discard data storage (and in cooperation with the SGDBI), and report back to the next planning group meeting (PGCCDBS).
- DIFRES be requested to develop a test version of VPABase, using a range of WG data, and report progress at the next PGCCDBS meeting.
- IPIMAR be requested to extend ISLDB as necessary to accommodate additional national data sets as discussed, and report progress at the next PGCCDBS meeting.
- Subject to approval at the next PGCCDBS meeting, these applications can be released for trials by stock coordinators, Working Groups, and the next Discards meeting as appropriate.
- ICES should establish and maintain standard codes for data to minimize problems associated with data exchange.
- Data exchange formats should be based on text files (delimited or XML), and variables should be clearly described and in agreement with the standard codes.

Recommendations concerning co-ordination:

- That ACFM establish PGCCDBS as an annual working group, and that PGCCDBS should meet in late February 2003.
- That meetings are convened to coordinate sampling to ensure adequate spatial and temporal sampling coverage.
- That WG species coordinators be asked to report to PGCCDBS the deficiencies in the commercial data contributing to stock assessment.

8 ACKNOWLEDGEMENT

The Planning Group participants thank the IPIMAR, Portugal for the invitation to meet in Lisbon and for providing excellent working facilities and service, together with beautiful weather.

9 REFERENCES

Commission Regulation (EC) No. 1639/2001 of 25 July 2001 establishing the minimum and extended Community programmes for the collection of data in the fisheries sector and laying down detailed rules for the application of the Council Regulation (EC) No 1543/2000.

Connolly, P., Trujillo, V., Pestana, G., Peireira, J.G., Flatman, S., Moguedet, P., Newton, A., & Motos, L. ICES CM 2001/P:2: Key issues concerning basic fisheries data for western EU waters.

DELIASS: Development of elasmobranch assessment. EU Study project 99/055.

EFAN: European Fish Ageing Network. EU FAIR- CT 96 1304.

EMAS: Evaluation of Market Sampling. EU Study project 98/075.

Improvement of Data Collection for Stock Assessment in ICES Sub-areas VI, VII, VIII, IX, and X. Study Contract 94/0130.

FIEFA: Providing a Framework to Improve the Assessment of the main Demersal and Pelagic Fisheries in Western Europe. EU Study project 95/013.

IBSSP 1: International Baltic Sea Sampling Programme I. EU Study project 96/002.

IBSSP 2: International Baltic Sea Sampling Programme II. EU Study project 98/024.

ICES CM 1999/D:2: Manual for the International Bottom Trawl Surveys (Revision VI).

Maxwell, D.L., Rackman, B.D., Warnes, S., Armstrong, M., Newton, A.W., & Officer, R. ICES CM 2001/P:11: A study of market sampling data, precision of age compositions and the effect of mis-ageing: ICES Division VIIa cod.

MIQES: The use of multivariate data for improving the quality of survey-based stock estimation in the North Sea. EU Study project 97-009.

O'Brien, et al. ICES CM 2001/P:13: The precision of international market sampling for North Sea plaice (*Pleuronectes platessa* L.) and its influence on stock assessment.

SAMFISH: Improving Sampling of Western and Southern European Atlantic Fisheries. EU Study project 97/059.

SEC 2001 1798. Commission Staff Working Paper. Fisheries Control in Member States. DENMARK.

SEC 2001 1799. Commission Staff Working Paper. Fisheries Control in Member States. BELGIUM.

SEC 2001 1811. Commission Staff Working Paper. Fisheries Control in Member States. FINLAND.

SEC 2001 1812. Commission Staff Working Paper. Fisheries Control in Member States. FRANCE.

SEC 2001 1813. Commission Staff Working Paper. Fisheries Control in Member States. GERMANY.

SEC 2001 1814. Commission Staff Working Paper. Fisheries Control in Member States. GREECE.

SEC 2001 1818. Commission Staff Working Paper. Fisheries Control in Member States. IRELAND.

SEC 2001 1819. Commission Staff Working Paper. Fisheries Control in Member States. ITALY.

SEC 2001 1820. Commission Staff Working Paper. Fisheries Control in Member States. NETHERLANDS.

SEC 2001 1821. Commission Staff Working Paper. Fisheries Control in Member States. PORTUGAL.

SEC 2001 1822. Commission Staff Working Paper. Fisheries Control in Member States. SPAIN.

SEC 2001 1823. Commission Staff Working Paper. Fisheries Control in Member States. SWEDEN.

SEC 2001 1824. Commission Staff Working Paper. Fisheries Control in Member States. UK.

Table 1. Sampling overview per species and area for 2002

LA = Recording of length and age information.

L = Recording of length information.

LA=Latvia, NE=Netherland, NO=Norway, PL=Poland, PO=Portugal, RU=Russia, SP=Spain, SW=Sweden, UK/EN=England, UK/SC=Scotland.

Species	Area/division	Country																		
		BE	DK	EE	FI	FO	FR	GE	IR	IS	LA	NE	NO	PL	PO	RU	SP	SW	UK/EN	UK/SC
Baltic ICES AREA III (excl. Skagerrak)																				
Cod	IIIa S		LA																LA	
Cod	IIIb-d		LA																LA	
Dab	IIIb-d																			
Flounder	IIIb-c																			
Flounder	IIIb-d			LA	LA															
Herring	IIIa S		LA																LA	
Herring	IIIb-c		LA																	
Herring	III d		LA	LA	LA														LA	
Nephrops (functional unit)	IIIa S		L																L	
Plaice	IIIa		LA																LA	
Plaice	IIIb-d		LA																	
Salmon	IIIb-d		LA	L	LA														LA	
Sole	IIIa S																			
Sprat	IIIa S		LA																LA	
Sprat	IIIb-d		LA	LA	LA														LA	
Turbot	IIIb-d																		LA	
Eel	IIIb-d																			
Haddock	IIIa S																			
Hake	IIIa S																			
Perch	III d			LA	LA															
Pike	III d																			
Pike-perch	III d			LA	LA															
Saithe	IIIa S																			
Sea trout	IIIb-d				LA															
Whitefish	III d				LA															
Whiting	IIIa S		LA																	
North Sea (Skagerrak) ICES AREA IIIa(north)																				
Blue Whiting	IIIa N		LA																L	
Cod	IIIa N		LA																LA	
Haddock	IIIa N		LA																LA	
Hake	IIIa N		LA																L	
Herring	IIIa N		LA																LA	
Mackerel	IIIa N		LA																LA	
Nephrops (functional unit)	IIIa N		L																L	
Norway pout	IIIa N		LA																L	
Pandalus	IIIa N		L																L	
Plaice	IIIa N		LA																L	
Saithe	IIIa N		LA																LA	
Sandeel	IIIa N		LA																L	
Small sharks	IIIa N																		L	
Sole	IIIa N		LA																	
Sprat	IIIa N		LA																LA	
Dab	IIIa N																		L	
Whiting	IIIa N		LA																L	

Table 1. Sampling overview per species and area for 2002

LA = Recording of length and age information.

L = Recording of length information.

LA=Latvia, NE=Netherland, NO=Norway, PL=Poland, PO=Portugal, RU=Russia, SP=Spain, SW=Sweden, UK/EN=England, UK/SC=Scotland.

Species	Area/division	Country																		
		BE	DK	EE	FI	FO	FR	GE	IR	IS	LA	NE	NO	PL	PO	RU	SP	SW	UK	UK
																			EN	SC
ICES AREA II																				
Atlanto-Scandian Herring	Ila, V		LA									LA	LA							LA
Cod	SA I,II												LA				LA			
Saithe	SA I,II												LA							
Shrimp	SA I,II												L							
Greenland Hal.	SA I,II												LA							
Haddock	SA I,II												LA					LA		
Redfish spp.	SA I,II												LA							
North Sea & Eastern Channel ICES AREAS IV, VIId																				
Anglerfish	IV, VIId		LA																	LA
Argentines	IV		L									LA	L							
Blue Ling	IV																			
Blue Mouth	IV																			
Blue whiting	IV, VIId		LA									LA	LA							L(A)
Brill	IV, VIId											LA								
Catfish	IV																			
Cod	IV, VIId		LA									LA	LA							LA
Cuckoo ray	IV, VIId											(L)								
Dab	IV, VIId											LA	L							
Deep water sharks	IV																			
Fork Beard	IV																			
Greenland Halibut	IV																			
Haddock	IV, VIId		LA										LA							LA
Hake	IV, VIId		LA																	L
Herring	IV, VIId		LA				LA					LA	LA							LA
Horse mackerel	IV, VIId		L									LA	LA							(L)
Lemon sole	IV, VIId		LA									LA								L
Ling	IV																			
Mackerel	IV, VIId		LA									LA	LA							LA
Megrim	IV, VIId																			LA
Mullet	IV, VIId																			
Nephrops (functional u	IV		L									L	L							L
Norway pout	IV		LA										LA							
Other Rays & Skates	IV, VIId											(L)								L
Pandalus	IV		L										L							
Plaice	IV		LA									LA	L							L
Plaice	VIId						LA													
Red Mullet	IV, VIId						L													
Redfish	V																			
Rough head Grenadier	IV																			
Saithe	IV, VIId		LA				LA						LA							LA
Salmon	VI																			
Sand eel	IV		LA										LA							LA
Scallops	VIId																			
Seabass	IV, VIId						L													
Small sharks	IV, VIId																			
Sole	IV		LA									LA	L							
Sole	VIId						LA													
Spotted ray	IV, VIId											(L)								
Sprat	IV, VIId		LA										LA							LA
Spur Dogs	IV, VIId																			L
Starry ray	IV, VIId											(L)								
Thornback ray	IV, VIId											(L)								
Turbot	IV, VIId		LA									LA	L							
Tusk	IV																			
Whiting	IV, VIId		LA				LA					LA	L							LA
Witch	IV																			

Table 1. Sampling overview per species and area for 2002

LA = Recording of length and age information.

L = Recording of length information.

LA=Latvia, NE=Netherland, NO=Norway, PL=Poland, PO=Portugal, RU=Russia, SP=Spain, SW=Sweden, UK/EN=England, UK/SC=Scotland.

Species	Area/division	Country																		
		BE	DK	EE	FI	FO	FR	GE	IR	IS	LA	NE	NO	PL	PO	RU	SP	SW	UK	UK
																			EN	SC
NE Atlantic & Western Channel ICES AREAS II, V, VI, VII (exc d) VIII IX X XII XIV																				
Anchovy	IXa																		LA	
Anchovy	VIII						LA												LA	
Anglerfish	Vb,VI,XII,XIV VII,VIIIabde						LA	LA											LA	LA
Anglerfish	VIIIc,IX,X													LA					LA	
Blue whiting	I-IX,XII,XIV							LA			LA	LA		LA					LA	L
Cod	Vla, Vb, VIIa, VIIb-k,VIII, XII,XIV						LA	LA												LA
Haddock	Vb,VI,XII,XIV							LA												LA
Haddock	Vla, Vb, VIIa, VII,VIII,XII,XIV						LA	LA												LA
Hake	IIIa,IV,VI,VII,VIIIab VIIIc,IXa						LA	LA						LA				LA		L
Herring	Vla, VIabcj							LA												LA
Horse mackerel	IIa,IVa,V,VI,VII,VIII,IX		L					LA			LA			LA				LA		(L)
Mackerel	II,IIIa,IV,V,VI,VII,VIII,IX		LA					LA			LA	LA		LA				LA		LA
Megrim	Vb,VI,XII,XIV, VII,VIIIabcde,IX						LA	LA						LA				LA		LA
Nephrops (functional u	Vb,VI, VIIa (not 33E2- E5), VIIb,c,d,e,j,k, VIIIabde						L	L										L		L
Nephrops (functional u	VIIIc,IX													L				L		
Plaice	VIIa, VIIfg						LA	LA												
Plaice	VIIe						LA													
Plaice	VIIbc, VIIhk, VIII						LA	LA												
Pollack	all areas																			
Rays and Skates	all areas						L	L										L		L
Redfish	Va,XII,XIV						L											LA		
Saithe	Vb,VI,XII,XIV						L													LA
Saithe	VII,VIII						L	LA												
Sardine	VIII,IX						LA											LA		
Sole	VIIa/ VIIfg/VIIIab						LA	LA			LA									
Sole	VIIe						LA													
Sole	VIIbc, VIIhk,IXa							LA												
Whiting	IX														L					
Whiting	Vb, VI, XII, XIV, VIIa, VIIb-k, VIII						LA	LA												LA
Witch	VI,VII							L												
Alfonsinos	all areas																	LA		
Argentines	all areas										LA	L								
Blue ling	all areas											L		LA						
Bluemouth rockfish	all areas													LA						
Capelin	XIV											LA								
Conger	all areas											LA	LA					LA		
Crab	all areas						L	L										L		L
Cukoo ray	all areas						L													L
Cuttlefish	all areas																	L		
Forkbears	all areas													LA						
Greenland halibut	Va,XII,XIV											LA								
Horse mackerel (ssp)	VIII,IX											LA	LA	LA						
Lemon sole	all areas							LA												L
Ling	all areas						L	LA				L								
Lobsters	all areas						L	L				L								L
Meagre	all areas						L													

Table 1. Sampling overview per species and area for 2002

		Country																		
Species	Area/division	BE	DK	EE	FI	FO	FR	GE	IR	IS	LA	NE	NO	PL	PO	RU	SP	SW	UK	UK
																			EN	SC
Mediterranean Horse Mackerel	VIII,IX																			
Octopus	all areas														LA		L			
Orange roughy	all areas								LA											L
Other Deepwater species	all areas																			
Razor	all areas								L											
Pouting	all areas														LA		LA			
Red mullet	all areas						L													
Roundnose Grenadier	all areas						LA													
Salmon	all areas																			
Scabbards	all areas														LA					
Scallops	all areas						L													LA
Seabass	all areas						L										LA			
Seabreams	all areas														LA		LA			
Shrimp	all areas														L		L			
Spanish mackerel	VIII,IX														LA					
Spurdog	all areas								LA											L
Squids	all areas														L		L			(L)
Wedge sole	all areas																			
Whelks	all areas								L											
Wreckfish	X														L					

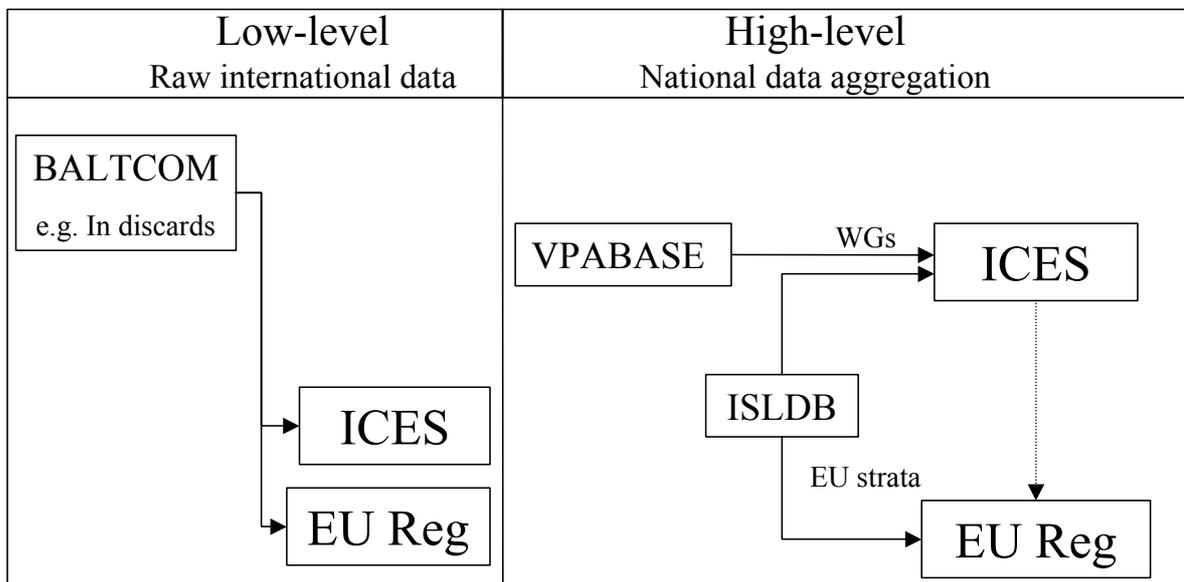


Figure 1. Alternative approaches based on available (but further developed) software

APPENDIX 1

CATCH INFORMATION SHEET

Information given by ICES Working group members.

Working Group name:

Chairman:

Species:

Stock:

Area/division:

Landings Which information source has been used to make up landing figures?

Country	Sales slips notes	Coverage level	Logbooks	Coverage level	Other sources (descripe which)
Belgium					
Denmark					
England					
Estonia					
Finland					
France					
Germany					
Ireland					
Latvia					
Lithuania					
Netherlands					
Norway					
Poland					
Portugal					
Russia					
Scotland					
Spain					
Sweden					

Discards Which information source has been used to estimate discards figures?

Country	Observer programmes	Coverage level	Logbooks	Coverage level	Other sources (descripe which)
Belgium					
Denmark					
England					
Estonia					
Finland					
France					
Germany					
Ireland					
Latvia					
Lithuania					
Netherlands					
Norway					
Poland					
Portugal					
Russia					
Scotland					
Spain					
Sweden					

APPENDIX 2

Descriptions of catch statistics

Working Group name: Chairman:

Species: Stock: Area/division: Year:

Landings

Describe the WGs view on the quality of the total international landing figures:

--

Are there problems with misreporting, “black” landings etc?
(if not known, please examine)

--

If the landing figure has de-/increased compared to last year figure, describe why.

--

Discards

Describe the WGs view on the quality of the total international discard figures:

--

If the discard figure has de-/increased compared to last year figure, describe why.

--

APPENDIX 3

a) Introduction to the BALTCOM database

A new web-based database called BALTCOM has been designed and implemented under the EU Study program International Baltic Sea Sampling Program II (IBSSP II). The database contains all raw data collected during the period 1 July 1995 up to present. Data can be uploaded to the database via a conversion program converting the data from the "exchange format" to XML. A data check procedure is performed during uploading of the data, and no data are accepted unless they pass this check. In connection with the database a data warehouse has been developed. The data warehouse offers the possibility to calculate all input tables of biological information (discards and landings) necessary to perform a standard assessment and to design tables on pivot basis similar to what is possible in EXCEL. Based on these tables plots and graphs can be made on an interactive basis. Furthermore, the data warehouse makes it possible to export data to a number of formats including EXCEL, SAS, and ASCII for additional analysis. All calculations can be made based on stratifications on fisheries. Access to the database can only be made through the National Co-ordinator using a password. National Coordinators can read, make analysis, export data, and after acceptance from involved countries, publish data from all countries. National Coordinators are only entitled to correct data from their own country. The database is situated on a server placed at DIFRES, Denmark. Key persons in DIFRES act as administrators and have access to all data at all levels. On-line help is available throughout the database. Each group of functionality in the database is constructed as a module, and a number of modules constitute the database. This organization facilitates an easy procedure when data extraction is carried out.

b) Introduction to the VPA-base

The VPABASE is a prototype for storage and processing of Working Group inputs for fish stock assessment. Input data are total catches by fleet and country combined with age distribution samples. The processing of data applies the available age distribution samples to convert the national catches into age compositions and sum up the national data to a total stock age composition, which is then the input to the VPA. The database will maintain records on the processing of data. The database is implemented in ACCES/EXCEL, using a user-friendly interface.

The database was developed as a part of the EU-funded project EMAS.

c) Introduction to the ISDLB database

The International Sampling Level DataBase (ISLDB) was designed to support the management of international sampling and to perform exploratory analysis of international sampling. It was initiated with the Study Contract 94/013 and further developed in the FIEFA and SAMFISH Projects.

d) Introduction to the DATRAS database

DATRAS (DATAbase Trawl Surveys) is an EU-funded project involving ICES, RIVO, IFERMER, and MARLAB in the re-construction and extension of existing survey databases held at ICES. The data will be derived from all coordinated surveys in the Baltic, the North Sea, and the western/southern divisions. The database will be held on a web-based Microsoft SQL-server and access will be at task levels. One level, which will only be accessible in ICES (database manager), will relate to import, updates, and maintenance, and one level for data extractions will be for all users and can be entered through the Internet. The new database will be functional by December 2003. A full description of the database can be found in the Working Document (Description of DATAbase Trawl Surveys, L. Larsen) presented to the DATRAS Steering Group, 10 January 2002.

APPENDIX 4

Presentation of ISLDB –A TOOL FOR SAMPLING (META)ANALYSIS by

Ernesto Jardim (ernesto@ipimar.pt)

IPIMAR, Av.Brasilia, 1400-006 Lisboa, Portugal

The development of the ISLDB has been financed by the Projects STUDY CONTRACT 94/013, FIEFA and SAMFISH. International Sampling Level DataBase (ISLDB) was developed in Access 2 and later migrated to Access 97, to:

“Analyse international sampling data, using parameters like the number of samples collected in a strata, the number of individuals measured and aged in the same strata or the sampled weight. Compare these parameters with the catches reported by species in a defined strata.” (Jardim, Artetxe & Gaudou; 2001)

The information included is the number of length samples collected and their weight, number of individuals measured, and number of individuals aged by institution, species, month, gear, and ICES Division. The Institutions contributing are the partners of the Samfish Project: MI (Ireland), MarLab (Scotland), CEFAS (England & Wales), IFREMER (France), AZTI (Spain), IEO (Spain), IPIMAR (Portugal), and DOP/UA (Portugal).

Problem:

- How is sampling performing in a stock perspective ?
- How can we assess the sampling levels and coverage on a stock perspective?
- How to combine sampling data collected by different sampling schemes ?

Solution:

- Use metadata (samples, individuals, etc.).
- Create a tool to store the necessary data, perform some routines (check, aggregate, etc.) and query data.
- Develop a protocol for sampling analysis.
- Compile information from the different partners sampling the same stocks.
- Analysing data from a stock perspective considering time/space/gear aggregation.

Major problems:

- Check the data submitted.
- Combine data from different sampling schemes.
- Combine data collected with different levels of aggregation.
- Different codes for gear, species, and areas.
- Make everything work.

The main reports:

- Sampling Level and Intensity (see database).

Other analysis:

- Sampling coverage (spacial analysis).
- Sampling over time (time-series analysis).

Future?

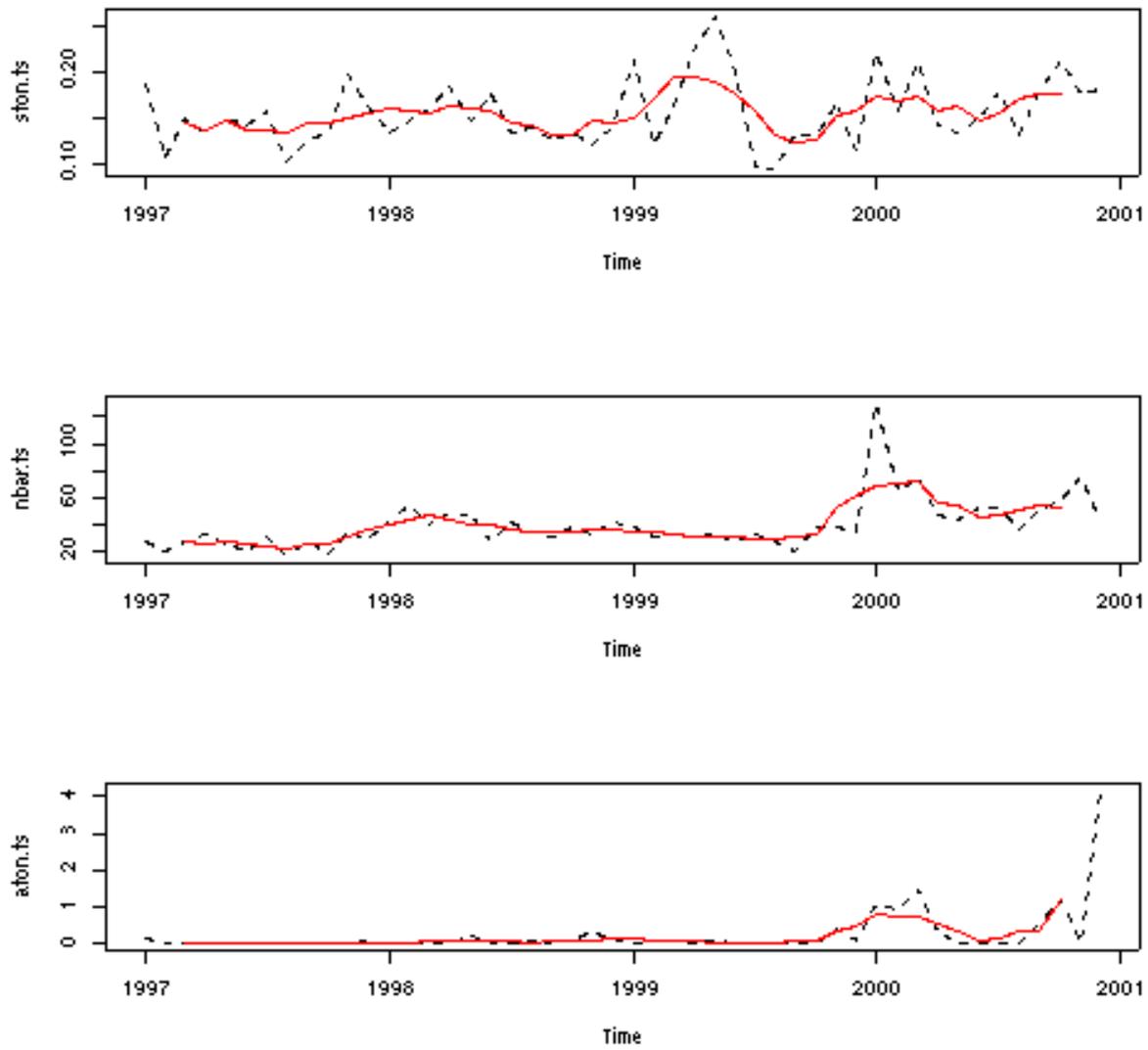
- International projects ended:
 - no contact between partners ?
 - no funding for further development ?

we must keep working !!

Where to go, now?

- Invite more partners to join the ISLDB group.
- Include more information (mean length and variance, mean age and variance).
- Migrate to a LAMP environment (Linux/Apache/MySQL/PHP) and make it available through the internet with security.

Figure 3. Sampling over time (time series analysis)



APPENDIX 5

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

ICES CM 2001 / P : 23

VPABASE: A PROTOTYPE OF DATABASE FOR STORAGE AND PROCESSING OF VPA-INPUT DATA

by

Per Sparre, Ole Folmer and Clara Ulrich

Danish Institute for Fisheries Research,
Charlottenlund Castle, DK 2920 Charlottenlund, Denmark.
email: pjs@dfu.min.dk, ofo@dfu.min.dk and clu@dfu.min.dk

Abstract

The paper describes a prototype of database for storage and processing of Working Group input for fish stock assessment. Input data are total catches by fleet and country combined with age distribution samples (when available). The processing of data applies the available age distribution samples to convert the national catches (landings and discards) into age compositions and sum the national data to total stock age composition, which is the input to the VPA. The database will maintain records on the processing of data, as well as records on misallocated and unreported catches, explaining differences between official data and so-called "WG-data". The database is implemented in ACCES/EXCEL, using a user-friendly interface, including a user's manual. Data may be entered through userforms or transferred from other databases through a file with a specified exchange format. The database was developed as a part of the EU funded project EMAS (Evaluation of MArket Sampling strategies for a number of commercially exploited stocks in the North Sea and development of procedures for consistent data storage and retrieval, CFP Study Project 98/075).

1	INTRODUCTION	27
2	INPUTS TO STOCK ASSESSMENT.....	28
2.1	Catch at age	29
2.2	Weight at age.....	29
3	ESTIMATION OF CATCH AND WEIGHT AT AGE.....	30
3.1	Biological sampling.....	30
3.2	Sampled vs. unsampled strata.....	30
3.3	Estimating catch at age in unsampled strata	31
3.4	SOP-correction	32
4	NATIONAL DATA AND ICES WORKING GROUP DATA.....	32
4.1	Official statistics and working group data	32
4.2	Discards and landings.....	33
4.3	Misreported and unallocated catch	34
5	(PROTOTYPE-) SOFTWARE	34
5.1	Tables and relations	35
5.2	User-interface	36
5.3	Manual data entry	36
5.4	Automatic data Input	37
5.5	Processing of data.....	37
5.6	Reports in Access format.....	38
5.7	Reports in Excel format.....	38
6	DATA-STORAGE AND DATA-EXCHANGE.....	39
7	DISCUSSION AND CONCLUSIONS.....	40
8	REFERENCES	41
9	LIST OF SYMBOLS	42
10	TABLES	43
11	FIGURES.....	49

List of Figures

Figure 1. Illustration of the gap-filling-process.....	49
Figure 2. The basic hierarchy of tables in VPABase.....	49
Figure 3. Data Tables and relations of the ACCESS implementation of VPABase.....	50
Figure 4. Data-tables, Look-Up-tables and relations of the ACCESS implementation of VPABase ,.....	51
Figure 5. Main menu of VPABase.	52
Figure 6. Data entry form of VPABase.....	53
Figure 7. Menu for data processing.....	54
Figure 8. Data entry form for the selection of age compositions to be used to fill-in unsampled catches.....	54
Figure 9. EXCEL worksheet used by the ICES Herring WG and the Mackerel, Horse mackerel, Sardine and Anchovy WG.....	55
Figure 10. Example of report in EXCEL format.....	56

1 Introduction

Biological advice on many commercially exploited stocks in the ICES area relies to a large extent on commercial catch-at-age. The quality of the biological advice is dependent on the quality of the underlying data. However, for most stocks there is hardly any insight into measures of uncertainty associated with the catch-at-age data. Furthermore, for the stocks exploited by various countries, the sampling and raising procedures to estimate national age-composition of catches might differ from one country to another. The methods used for combining national catch statistics into international catch at age are not uniform. The historical disaggregated national catch data are not always accessible, and thus the consistency of data series is not warranted. Thus the whole procedure of estimating international catch at age would gain in efficiency and transparency if a uniform method of aggregation was developed and applied, and if all data used in the estimation were gathered and centralised in one single database.

The objective of this paper is to present a prototype of such a database for the storage and the processing of ICES Working Group input for fish stock assessment. Input data are total catches by fleet (and country), area and time period combined with age distribution samples (when available). It was developed as a part of the EU funded project EMAS (Evaluation of Market Sampling strategies for a number of commercially exploited stocks in the North Sea and development of procedures for consistent data storage and retrieval, CFP Study Project 98/075; Pastoors *et al.*, 2001). This project intended to evaluate the adequacy of the international market sampling effort for a limited number of commercially exploited fish stocks in the ICES area, to quantify the uncertainties in estimated catch in numbers and mean weight at age, to advise on appropriate sampling levels and methods and to define protocols and develop prototype software to store and aggregate national sampling data.

The data processing uses the available age distribution samples to convert the national catches (landings and discards) into age compositions, and sums the national data to total stock age composition, which are the main input to the VPA (Virtual Population Analysis). The processing of data implies filling in gaps in data, caused by some countries not supplying full sets of input data. The gaps are filled in by extrapolating samples data to non-sampled catches data (Figure 1).

The database is a relational database implemented in ACCESS, chosen for its simplicity of use and because it is a standard software commonly used in fisheries institutes. Although ACCESS might not be designed for storing and manipulating large-scale databases, it was considered a suitable tool for the present application. It allows having a user-friendly interface. Data may be entered through user-forms or transferred from other databases through a file with a specified exchange format. Reports from the database are made in EXCEL sheets. A user's manual is available.

A number of software tools were already developed for data storage and combination of national catch information to the international level:

- (a) In the ICES Herring Assessment Working Group for the Area South of 62° N (HAWG) national data are submitted on agreed spreadsheet forms. The software called SALLOC is used to raise national data to the international level (Patterson 1998). This Fortran program provides a log listing the actual fill-ins used to cover all landings with missing biological information, but this list is not given in the WG report. Also, no list is available in the WG report giving the extrapolations made of each national data set before submitting the data to ICES. The Working Group provides a well-documented overview of the sampling coverage from each country in the report.
- (b) The Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK) keeps a database system to store and aggregate data on the flatfish species in the North Sea. The database system (FISHBASE) consists of a user interface programmed in Pascal and a specific datafile format. Data are stored in separate datafiles by country, gear, area, sex (if available) and year. Different levels of aggregations may be used in the datafiles ranging from files, which have catch-numbers, weights at age and lengths at age by quarter, sex, gear, area and country to files, which only have total landings, by country and year. Raising is performed from the lowest level aggregation upwards. The strong points of Fishbase are that it forces the user to correctly specify the input data and that data can be exchanged in separate datafiles, so that national institutes can submit their data in those files. Also it allows a historical access to the underlying data and other arrangements may be implemented to raise the data. Negative points are that no manual is available, that the program does not allow combination of areas, that the input format is very difficult to generate with any other software than Fishbase itself and that there is no specification of reallocation rules.

- (c) A series of basic programs are used to collect national age compositions and weights at age for roundfish stocks in the North Sea and west of Scotland. The programs were developed in the 1980s in the Marine Laboratory (Aberdeen) and are operated from the dos-prompt by invoking QBASIC. All data is stored in a common format. An exchange format is used for exchange of data. Strong points of the Aberdeen programs are that a rigid file structure is defined and that the raising procedure is carried through to generating the input files for stock assessment. Negative points are that there is a whole suite of programs, which should be used, in a specific order and that there are no explicit reallocation rules.

In the software presented here, the raising of catch-at-age distributions to non-sampled data is done automatically, with the same algorithm as in Patterson (1998). But improvements of the current work are threefold: Firstly, this algorithm is combined directly with the database, simplifying the manipulation and processing of data. Secondly, it is aimed to be a more generic and flexible tool that could be used for a number of different fisheries. Thirdly, it includes a higher level of data-disaggregation to the fleet level. This disaggregation is useful for improving the estimation of total catches at age, by taking into account the differences of exploitation pattern arising from differences in fishing practices.

The database is thus intended to output the inputs to the VPA. We consider here only the usual single-species VPA used in most ICES Working Groups. Multi-species VPA, although used for the assessment of some specific stocks, is not considered at this stage of the database elaboration. We do not consider the VPA tuning process, which involves further specifications of time series of effort and catches for some commercial and surveys tuning fleets. However, it is worth noting that as the database is implemented at the fleet level so that future improvements could easily be provided to store tuning data as well in the database.

In this paper we firstly summarise briefly the main inputs to be used in the VPA, and how they can be calculated from sampling data. We then summarise how and why these inputs are modified during working groups before being used in assessment. This aims to present the full set of parameters that should be included in the database. The database itself is presented afterwards.

2 INPUTS TO STOCK ASSESSMENT

The VPA compute stock numbers and fishing mortality coefficients by age and year, from data of total catches in number and estimates of natural mortality coefficients, by age and year (although natural mortality is most often considered as constant over years). Biomass is thus calculated by multiplying stocks numbers by weight at age.

To facilitate the notation, we introduce the “*Stratum*”. By a *stratum* is meant a combination of a stock St , a country Ct , a fleet Fl , an area Ar and a quarter q in a year y . We use the symbol “ Ω ” for a stratum:

$$\Omega = (St, y, Ct, Fl, Ar, q).$$

The symbol “ S ”, represent a set of strata: $S = \{\Omega_1, \Omega_2, \dots, \Omega_m\}$, where m is the number of elements in the set S . Appendix 1 contains a complete list of symbols used in the formal description.

Catch at age

The input to single species VPA without tuning, are the numbers caught (landed and discarded) of stock St by age group a and during quarter q in year y , from the total area of distribution of the stock, caught by all fleets exploiting the stock.

$$C_{VPA}(St, y, q, a)$$

We use here the quarter as the time step by default, as it is commonly used in the ICES area (although assessments are afterwards conducted on an annual basis). Any other division of the year, for which data are available, could be used. The age index is comprised in the range $[a_{Min}(St), a_{Max}(St)]$, where $a_{Min}(St) (\geq 0)$ is the youngest age group considered in the VPA and $a_{Max}(St)$ is the oldest age group considered in the VPA. The oldest age group may or may not be a plus-group.

The catch is taken by a number of fleets in a number of sub-areas of the total stock area, so that

$$C_{VPA}(St, y, q, a) = \sum_{Ct=1}^{Ct_{Max}} \sum_{Fl=1}^{Fl_{Max}(Ct)} \sum_{Ar=1}^{Ar_{Max}} Catch(St, y, q, a, Ct, Fl, Ar) = \sum_{Ct=1}^{Ct_{Max}} \sum_{Fl=1}^{Fl_{Max}(Ct)} \sum_{Ar=1}^{Ar_{Max}} Catch(\Omega, a) \quad (1)$$

where $Catch(\Omega, a) = Catch$ (Landings + discards) in numbers of age group a in a stratum Ω .

$Fl = 1, 2, \dots, Fl_{Max}(Ct)$, $Ar = 1, 2, \dots, Ar_{Max}$, where $Fl_{Max}(Ct) =$ Number of fleets in country Ct and $Ar_{Max} =$ Number of areas.

The area can be ICES sub-division, ICES division or ICES area. In a later version of the database, the area concept could be generalized to include ICES rectangles, and/or specific areas defined by working groups. Fleets are given homogeneous subsets of boats, whose definition is made externally by the Working Group (e.g. by fishing gear and/or boat length class...)

In general the dot “•” is used to indicate summation over one index, e.g. when data are not disaggregated by quarter but are available only on an annual basis :

$$Catch(St, y, \bullet, a, Ct, Fl, Ar) = \sum_{q=1}^4 Catch(St, y, q, a, Ct, Fl, Ar)$$

2.1 Weight at age

Estimated stock numbers are converted into stock and spawning stock biomasses by using weight and maturity coefficients by age and year. Predictions of yield are made from mean fishing and natural mortality at age and mean weight at age. The weight at age is thus a major parameter of the assessment procedure. It relates to a specific time, e.g. the beginning, middle, or mean over the time period. In subsequent equations the weight at age is noted :

$W_C(St, y, q, a, Ct, Fl, Ar) = W_C(\Omega, a) =$ Mean body weight in Catch of age group a in a stratum.

Number of fish caught is calculated as :

$$Catch(\Omega, a) = \frac{Yield(\Omega, a)}{W_C(\Omega, a)} \quad (2)$$

Where $Yield(\Omega, a)$ is the catch in weight of stock St and age a , in a stratum. The total yield $Yield(\Omega)$ must be available for all strata, to make the following estimation feasible..

3 ESTIMATION OF CATCH AND WEIGHT AT AGE

3.1 Biological sampling

Catch and weight at age are estimated from biological sampling data, which provide detailed information on length and/or age composition of the landings in a stratum. Data include the number of otoliths read and the number of fish measured by age. The associated coefficients of variation may be available. Sampling data by age and stratum in the database are thus :

$na(\Omega, a)$ = Number of age measurements (otoliths);

$nl(\Omega, a)$ = number of length measurements;

$CV_C(\Omega, a)$ = Coefficient of variation of the estimate of Catch(Ω, a);

$CV_{wC}(\Omega, a)$ = Coefficient of variation of the estimate of $w_C(\Omega, a)$

3.2 Sampled vs. unsampled strata

A *sampled stratum* is a stratum in which biological sampling was conducted. The catch numbers at age and the weights at age could be directly estimated. The total yield is also available. An *un-sampled stratum* is a stratum where only the total yield is known, and no sampling data are available. The estimation of age composition and weight-at-age for un-sampled strata has thus to be done using information from the sampled strata.

The set of all strata may be split into two, the sampled strata ($S^{sampled}$) and the un-sampled strata, ($S^{Un-sampled}$). The number of sampled and unsampled strata are noted m_S and m_U respectively. A sampled and unsampled stratum is noted Ω^S and Ω^U respectively. Thus, the set of all strata, can be partitioned as (“ \cup ” indicates the union of two sets):

$$S^{All} = S^{sampled} \cup S^{Un-sampled} \quad S^{Sampled} = \left\{ \Omega_1^S, \Omega_2^S, \dots, \Omega_{m_S}^S \right\} \quad S^{Un-Sampled} = \left\{ \Omega_1^U, \Omega_2^U, \dots, \Omega_{m_U}^U \right\}$$

In a sampled stratum Ω^S , the total number caught of all age groups is:

$$Catch(\Omega^S, \bullet) = \sum_{a=a_{Min}(st)}^{a_{Max}(st)} Catch(\Omega^S, a) \quad (3)$$

The proportion of each age group in the total catch number in the stratum is:

$$Pr_{Catch}(\Omega^S, a) = Catch(\Omega^S, a) / Catch(\Omega^S, \bullet) \quad (4)$$

The mean weight of all fish in the catch is

$$W_C^{Mean}(\Omega^S) = \sum_{a=a_{Min}(st)}^{a_{Max}(st)} Pr_{Catch}(\Omega^S, a) * W_C(\Omega^S, a) \quad (5)$$

The computed total weight of the catch (SOP, Sum Of Products) is:

$$SOP_{Catch}(\Omega^S) = \sum_{a=a_{Min}(st)}^{a_{Max}(st)} Catch(\Omega^S, a) * W_C(\Omega^S, a) = W_C^{Mean}(\Omega^S) * Catch(\Omega^S, \bullet) \quad (6)$$

3.3 Estimating catch at age in unsampled strata

The age composition within an unsampled stratum can be estimated using the age distribution of other strata, which can be assumed to have a comparable pattern.

Let Ω_i^U be an un-sampled stratum, and let Ω_j^S be a sampled stratum. If we assume that weight-at-age and relative catch-at-age are the same for Ω_i^U and Ω_j^S , then we can estimate the total number caught in the unsampled stratum as:

$$\text{Catch}(\Omega_i^U, \bullet) = \text{Yield}(\Omega_i^U) / W_C^{\text{Mean}}(\Omega_j^S) \quad (7)$$

and

$$\text{Catch}(\Omega_i^U, a) = \text{Catch}(\Omega_i^U, \bullet) * \text{Pr}_{\text{catch}}(\Omega_j^S, a) \quad (8)$$

It may also be preferable sometimes to use a subset of various sampled strata to estimate catch distribution in an unsampled stratum, instead of only one as described in the previous equations. The estimated distribution is thus a weighted mean of the age distributions in sampled strata. The weighting coefficients might be uniform (equal weight for all strata), or linked to the importance of each stratum (total yield, number of fish measured/aged, inverse of variance...etc...).

The subset of sampled strata used for the estimation is denoted S_k ($S_k \subset S^{\text{Sampled}}$) and has k elements, $S_k = \{\Omega_1^S, \Omega_2^S, \dots, \Omega_k^S\}$.

we define: $\text{SW}_{Mt}(S_k, a) = \{v_{Mt}(\Omega_1^S, a), v_{Mt}(\Omega_2^S, a), \dots, v_{Mt}(\Omega_k^S, a)\}$

where SW is a set of weights ($v_{Mt}(\Omega, a)$) associated with the set of sampled strata and age group a , selected by a weighting method Mt .

The weighted mean proportion at age in the unsampled strata is then:

$$\text{Pr}_{\text{Catch}}^{\text{Mean}}(\Omega_i^U, a, S_k, Mt) = \frac{\sum_{\Omega^S \in S_k} \text{Pr}_{\text{Catch}}(\Omega^S, a) * v_{Mt}(\Omega^S, a)}{\sum_{a=a_{\text{Min}}(St)}^{a_{\text{Max}}(St)} \sum_{\Omega^S \in S_k} \text{Pr}_{\text{Catch}}(\Omega^S, a) * v_{Mt}(\Omega^S, a)} \quad (9)$$

The calculation of a weighted mean proportion at age is only possible when the weighting factors are available for all strata in S_k . The methods $Mt = \text{equal weight}$ (all $v = 1.0$) or $Mt = \text{yield}$ are applicable for any set of strata. The method $Mt = \text{number of otoliths}$ requires having an associated number of otolith measurements for all strata.

Note that the procedure to calculate weighted mean proportions at age may also be applied to sampled strata. The otolith samples taken from a catch stratum are usually used to estimate the age composition of that stratum. However, if the sample is very small or for some other reason suspicious, it may be better to use alternative otolith samples, or to combine the suspicious sample with samples of other strata.

The weighted mean body weight in the catch is then :

$$W_C^{\text{Mean}}(\Omega_i^U, a, S_k, Mt) = \frac{\sum_{\Omega^S \in S_k} W_C(\Omega^S, a) * v_{Mt}(\Omega^S, a)}{\sum_{\Omega^S \in S_k} v_{Mt}(\Omega^S, a)} \quad (10)$$

Then we can calculate the total number of individuals caught (using Eq. 7) and then the number of individuals per age group (using Eq. 8).

A list of possible weighting options for the calculations of proportions at age and mean weights in the catch is presented in **Table 1**. The list of options for weighting factors may be extended in future versions of VPABase. Obvious extensions of the list includes characteristics of ALKs (Age length keys), such as number of length measurements and number of ALKs.

3.4 SOP-correction

The SOP is the sum of the products of catch numbers at age times the weights at age:

$$SOP_{Catch}(\Omega) = \sum_{a=a_{Min}(St)}^{a_{Max}(St)} Catch(\Omega, a) * w_C(\Omega, a) \quad (11)$$

Ideally, the SOP should be equal to the Yield(Ω) which was given as input. If that is not the case, the equality can be achieved by applying a “SOP-correction”, either to the catch or to the weight

$$Yield(\Omega) = \sum_{a=a_{Min}(St)}^{a_{Max}(St)} Catch(\Omega, a) * Correction(\Omega) * W_C(\Omega, a) \quad (12)$$

where $Correction(\Omega) = \frac{Yield(\Omega)}{SOP_{Catch}(\Omega)}$

The SOP correction was discussed by Lewy & Lassen (1997), who concluded that the SOP is an unbiased estimate of the total catch, and that SOP corrections should not be made. We consider four options in the case of SOP discrepancies:

- 1) Correct the catch-at-age,
- 2) Correct the weight-at-age,
- 3) Correct the input Yield (replace the value of input Yield by the value of SOP),
- 4) Do nothing.

According to Lewy and Lassen, option 3 is recommended. In the following, it is assumed that any SOP-problem has been resolved before the data are entered into the database. If however, there is a discrepancy, the value Yield will be automatically changed to the SOP (Option 3).

4 NATIONAL DATA AND ICES WORKING GROUP DATA

The annual yield by country (summed over fleets) and ICES area are reported by member governments to ICES as the Official statistics. These statistics are however sometimes modified by members of assessment working groups. This section discusses the discrepancies between official statistics and the so-called “working group data”, as all the data included in the “working group data” have to be included in the database.

4.1 Official statistics and working group data

Let Landings $W_{Official}(St, Ct, Ar)$ be the official statistics, i.e. the weight of the annual landings of stock St by country Ct in ICES area (sub-area or division) Ar . Members of ICES working groups deliver the so-called *working group data*, which may be more disaggregated than the official data and which may also be different. The question is whether the official data are consistent with the working group data, that is

$$\text{Landings}W_{\text{Official}}(\text{St}, \text{Ct}, \text{Ar}) = ? = \sum_{\text{Fl}=1}^{\text{Fl}_{\text{Max}}(\text{St})} \sum_{\text{Ar} \in \text{ICES-Div.}} \sum_{\text{q}=1}^4 \text{Yield}(\text{St}, \text{Fl}, \text{Ar}, \text{q}) \quad (13)$$

Reasons for discrepancies might include the non-integration of discards data, as well as corrections made on misreported data.

The following fields are used by the Working Groups (ICES, 2000), and are integrated in the VPABase :

Official Catch	Catches as reported by the official statistics to ICES
Unallocated Catch	Adjustments to the official catches made for any special knowledge about the fishery, such as under- or over-reporting for which there is firm external evidence. (can be negative)
Area misreported Catch	Used only to adjust official catches which have been reported from the wrong area. (can be negative). For any country the sum of all the area misreported catches should be zero
Discarded Catch	Catch which is discarded
WG Catch	The sum of the 4 categories above

4.2 Discards and landings

Estimated catches (in numbers) and yield (in weight) is composed of landings and discards:

$$\text{Catch}(\Omega, a) = \text{Landings}N(\Omega, a) + \text{Discard}N(\Omega, a),$$

$$\text{Yield}(\Omega) = \text{Landings}W(\Omega) + \text{Discard}W(\Omega)$$

The weight at age of discards and landings are usually different, the discard weight-age-age being smaller than the landings weight-at-age:

$W_D(\Omega, a)$ = average body weight of discards.

$W_L(\Omega, a)$ = average body weight of landings.

The body weights relate to the body weight-at-age in the catch by:

$$W_C(\Omega, a) = \frac{W_L(\Omega, a) * \text{Landings}N(\Omega, a) + w_D(\Omega, a) * \text{Discards}N(\Omega, a)}{\text{Landings}N(\Omega, a) + \text{Discards}N(\Omega, a)} \quad (14)$$

whereby:

$$\text{Landings}W(\Omega) = \sum_{a=a_{\text{Min}}(\text{St})}^{a_{\text{Max}}(\text{St})} w_L(\Omega, a) * \text{Landings}N(\Omega, a)$$

and:

$$DiscardsW(\Omega) = \sum_{a=a_{Min}(St)}^{a_{Max}(St)} w_D(\Omega, a) * DiscardsN(\Omega, a)$$

4.3 Misreported and unallocated catch

When including the terms for unallocated and misreported landings as defined in the text table above, the following equations should always hold.

$$CatchW_{WG}(St, Ct, Ar) = LandingsW_{Official}(St, Ct, Ar) + UnallocatedW(St, Ct, Ar) + MisreportedW(St, Ct, Ar) + DiscardW(St, Ct, Ar) \quad (15)$$

where

$$\sum_{Ar \in ICES-Div.} Misreported(St, Ct, Ar) = 0$$

and

$$CatchW_{WG}(St, Ct, Ar) = \sum_{Fl \in Fleets(Ct)} \sum_{Ar \in ICES-Div.} \sum_{q=1}^4 Yield(St, Fl, Ar, q) \quad (16)$$

This equation must be met by all input data of each country. If it is not met, the Yield(St, Fl, Ar, q) will be corrected by the database system, using the country and area specific correction factor:

$$CorrectionWG(St, Ct, Ar) = \frac{LandingsW_{Working\ Group}(St, Ct, Ar)}{\sum_{Fl \in Fleets(Ct)} \sum_{Ar \in ICES-Div.} \sum_{q=1}^4 Yield_{BeforeCorrection}(St, Fl, Ar, q)} \quad (17)$$

where $Yield_{BeforeCorrection}(St, Fl, Ar, q)$ and $Yield_{AfterCorrection}(St, Fl, Ar, q)$ stand for Yield before and after correction, respectively.

$$Yield_{AfterCorrection}(St, Fl, Ar, q) = Yield_{BeforeCorrection}(St, Fl, Ar, q) * CorrectionWG(St, Ct, Ar) \quad (18)$$

However, basically, working group members should supply the data, so that no correction would be necessary. The program will always compute the correction factor, and the user should check the value. The database will keep the original value (before correction), for the purpose of documentation.

5 (PROTOTYPE-) SOFTWARE

A prototype database system was developed under the working title of “VPABase”. The prototype version of VPABase was implemented in Microsoft ACCES and EXCEL (version 2000).

The term “Prototype” means a “Working model in the early stage of the design process”. The prototype does run and executes some of the tasks specified in the foregoing sections. The prototype, however, is far from the stage where it is safe to distribute to ICES WGs. It has not yet been conscientiously tested, and so far no feedback has been received from the potential users, the ICES WGs. Thus, an unknown number of man-months of development, testing and consultations with the clients are still pending, before public distribution should be made.

A high degree of user-friendliness is aimed at in the design of user-interface and the output-options of VPABase. This feature has been attempted from the early stage of the development. Primary versions of a user’s manual and a reference manual of the VPABase has been prepared. This Section summarizes the two manuals.

The database can read input data from Excel-formats and also produce reports in Excel format. The example of EXCEL work sheet used in the present version of VPABase, has been adopted from the spreadsheets used by the Herring WG, and the Mackerel, Horse Mackerel, Sardine and Anchovy WG (ICES, 2000).

Input and output can also be handled out with plain ASCII text files (comma separated).

The VPABase is constructed as a relational database according to the standard database-design rules (normalization rules).

5.1 Tables and relations

The tables and the relations of VPABase consist of: “Input-data-tables”, “Look-up tables” and “Auxiliary tables”. The look-up tables are used to select standard input values, whenever possible, as that will reduce the risk of mistakes on data entry.

The principal data-tables of VPABase, contains the numbers at age and weight at-age distributions and the yields (weight of catches). The three principal data-tables of VPABase are:

1. Country record
2. Fleet record
3. Age composition record

In addition there is the auxiliary data tables, containing information on the data-processing:

4. All combinations of catch and samples for estimation of age composition.
5. Selected samples for estimation of age composition.

The tables are linked in a hierarchy of “one-to-many” relationships. Table 2 shows the fields in the input data-tables and the auxiliary tables. The first fields in the tables, e.g. “Key_Country_Catch” in the table “T_Country_Catch”) are the keys linking the tables in the hierarchy.

Table 2 indicates the relationships between the tables, which is also illustrated by Figure 3, which by arrows indicates the relations of the three principal data-tables, and the auxiliary tables.

Figure 4 shows the complete set of look-up tables together with the three principal input data tables.

A complete description of all fields, keys and relations of VPABase is rather voluminous. The reader interested in further details is referred to the reference manual of VPABase.

The auxiliary tables “T_Use_of_Samples” and “T_Use_Of_samples_Saved” contain the input, which specifies the processing of data. It contains information on which sampled records were used to estimate age-compositions of the un-

sampled records. They also contains information on the methodology of data processing as well as the time of (last) processing. This part of the database is perhaps the most complicated part of it, as the number of choices for combination of sample and processing method is usually large. A major objective of VPABase is to store the information on data-processing in an unambiguous way.

5.2 User-interface

The purpose of the user interface is to guide the user through three steps:

1. Input data;
2. Make decisions on estimation of age-compositions for un-sampled strata by selecting age-compositions among samples strata;
3. Generate output tables.

The user interface implemented through user-forms that are of two types:

1. Menus (selection of options)
2. Data entry form;

The five first items of the main menu (Figure 5) guides the user through the different steps to produce the aggregated input data for VPA. Options 1 and 2 are alternative options, namely (1) to “Enter data from the keyboard” (see next section) or (2) “Enter data from an EXCEL worksheet” (see Section 5.4). Item 1 in the main menu ends with “.....” which indicates that this option leads to another user-form (see Figure 6). Item 3 activates the user-form for data processing. Items 5 and 6 deal with the reports, that is the output. Output can be in ACCESS-format (Item 4) or in EXCEL-format (Item 5).

In later versions of VPABase, the main menu will contain an option for entry of tuning data (survey data, commercial fishery effort data or commercial fishery CPUE).

5.3 Manual data entry

The first item in the main menu opens a data input form (Figure 6) where data can be filled in manually. One country record might contain several fleet records and one fleet record might contain several sampling records. The form in Figure 6 is actually composed of three forms, which correspond to levels of the hierarchy of tables (see Figure 6 or Table 2)

First level – Annual catch by country: The top part of the data-entry-form is related to the country catch record and the record selector in the bottom of the form tells the user which country catch record is current. The record selector is the field called “Record:” and the arrow buttons gives the user a way to go the first record, move one step back, move one step forward, to the last record and to a new record.

Second level - Catch by fleet and time period: The second part of the data-entry-form contains the fleet record, with fleet related information. The fleet information in question relates to the country record, in which the fleet record is embedded. Ideally, the sum of the fleet records over time periods of a country should sum up to the value of the country record.

Third level – Age composition sample: The third part of the form contains the age composition data (samples), relating to the fleet record in which it is embedded. The sample is automatically given a unique “sample number”. This number cannot be changed by the user and is only used to identify a sample for the later processing of samples.

The data entered through look-up tables, appear as “list-boxes” in the user-form. When clicking on the right hand side arrow-button, a list of options is displayed, from which you can select the input. (The example given here is taken from Figure 6, showing the entry of “Stock”). Only input from the list is accepted. Only users with special “privileges” can change the content of the list box. This approach should reduce the risk of mistyping and ambiguities in data.

In later versions of VPABase, the data entry form will contain a number of additional facilities, such as production of graphs and tables as well as facilities for “finding” records and for deleting records.

Later versions of VPABase will also be able to handle catch data by statistical rectangle.

Entry of tuning data in future versions of VPABase, will probably be done in a separate userform, allowing for entry of catch and effort or CPUE.

5.4 Automatic data Input

The second item on the main menu lets the user import data from a predefined EXCEL format or a text file (see example of EXCEL input in Figure 9). The present version of VPABase can read the EXCEL sheets currently used by the Herring WG and the Mackerel, Horse mackerel, Sardine and Anchovy WG (ICES, 2000).

In later version of VPABase, input data may also be transferred in ACCESS table formats from one replicate of VPABase to the master version of VPABase, which operation should be managed by an international data-coordinator.

5.5 Processing of data

The third item in the main-menu, “Select samples from un-sampled catch, and process data” opens a submenu (Figure 7) for data processing.

If a fleet-record has an age composition sample attached to it (is a “sampled catch”), the user does not need to take further action as far as processing is concerned. In that case the VPABase will make the estimation of the age composition of the total catch from that sample. The user need only take action in the two cases:

- 1) When no age composition samples is available,
- 2) When the user wants to use additional age composition samples, for a sampled catch.

The action to take is to select one or more samples to represent the unsampled catch. If more than one sample is selected, the user will also have to specify which method should be used to pool the selected samples (how the weighted average should be calculated).

Before any actual selection of samples can be made, the user has to create a table to hold the tables of potential selections (the data-table “T_Use_Of_Samples, See Figure 3).

After having created the table of potetial samples the user can move to the next item on the processing-menu and start making selection for un-sampled strata (Item 2: “Select samples for unsampled catch”). The userform for selction of samples is shown in **Figure 8**.

The upper left part of this form contains a number of fields (headed “Catch record”) showing the particulars of the catch record to be processed. These fields are shown only for information. The fields are “locked”, so that the data cannot be entered or modified through this form.

There is a field in the centre of the form, which in the example of **Figure 8** contains the information “No age distribution for this catch” (written in red). This means that the catch of the current record is un-sampled, and you need to select samples, for estimation of age composition. If there is an age composition sample, the field will be empty.

The form has a check field in the right column to indicate which sample is selected for a given catch presented in the top portion of the form. The user can change the selection by removing or adding check marks on the form.

In the hypothetical example of **Figure 8**, two samples (no. 138 and 139) originating from “United States industrial fishery” is used to represent Danish industrial fishery.

In the upper right hand corner of the form it is possible to change weighting method, for pooling of samples. In the example of **Figure 8**, the samples have been weighted by the “Sampled catch”. The weighting methods only applies if more than one sample is selected for a given catch. The default method is to give equal weight to each sample, but if additional information, i.e. numbers of otoliths, is present this can be used when the samples are combined.

After the selection has been made, the user moves back to the processing menu where she/he can save the selection to a permanent table (data table “T_Use_Of_Samples_Saved”, see Figure 3) so the selection scheme is stored for later documentation of the process.

Default for the creation of the table is that if a catch record has a sample, this sample will be used for this catch (but there is the option not to use the default: Option 2 above).

5.6 Reports in Access format

The ACCESS system can create reports (or “output-tables) on its own. The ACCESS-format tables, however, are not very flexible with respect of additional processing, and change of style (see example in Table 4). Tables in EXCEL format is believed to be more attractive for members of ICES working groups.

Reports in Access format are inventory reports of the database. They contain information on the number of records and other features of the input data and the national sampling programs. Prototypes of the inventory reports are given in Table 4 (country records) and Table 5 (fleet records).

5.7 Reports in Excel format

These are the standard tables suggested for (potential) inclusion in the ICES Assessment working group reports. Ideally, the assessment-software should be able to take its input directly from these tables. The present proto-type version of VPABase, actually contains very few table. Most of the tables listed here cannot be produced by VPABase as the system is now. The authors of the present paper, await the reaction and the wishes from the possible “costumers” of VPABase, before they will start the work of designing output tables. So what follows is what the present authors expect will become the requirements from ICES WGs, but this still has to be tested by consultations.

The Excel reports of VPABase can be grouped as:

- 1) Input data, as provided by WG-members and governments.
 - a. Country records (before processing)
 - b. Fleet records (before processing)
- 2) Processing: Tables showing the methods and input
- 3) Results of processing
 - a. Country records (after processing)
 - b. Fleet records (after processing)
- 4) VPA input (Landings, Discards and Catch)

In principle, the database can display the data with any sorting or any grouping on the user choice. Some tables, however, are expected to be in demand and for these there will be pre-prepared routines to produce them.

These reports will comprise the traditional tables for official statistics of ICES working group reports, as well as the “non-official” data. The example in Table 6 shows a table for all country-records of a given ICES area. The system will be able to rearrange the tables on the users choice, for example to give a report by country with each row representing an ICES area.

Table 7 shows a proto-type of fleet-record table. There may or may not be age/weight-composition data available for a stratum. If age/weight-composition data are not available, then total catch (or landings) may be available, in which case only the last line in the table will be filled. The fleet records have unique samples numbers (*Sample No. y*), which are used for identification of the fleet-records when used for estimation of non-sampled strata.

An important requirement for the database system is to document the age composition estimation procedure. To document the process of using age compositions from other strata, means to record from where the “borrowed” information originates, and which method of weighting was applied. Table 7 shows a report which for each fleet record contains a log of its processing. Each row in the table represents a stratum. The first column gives the unique index of the stratum. The following 5 columns give the particulars of the sampled stratum. The columns *Landing Samp.* and *Discard Samp.* indicates if these catches were covered by the sampling program. The following columns contain the selection of “borrowed” samples used for estimation of landings- and discards age compositions. The first line contains the methods of weighting the data, if average values of several strata were used.

For the processed fleet records, the output will be a table similar to the output table for the sampled strata (Table 6). The table will contain information on the processing, extracted from Table 8. The results for the country, is either the sum over fleets or the estimation based on “borrowed” catch-distributions (Table 10).

The input to the stock assessment process, is the summation over country-records (Table 10). The only EXCEL report produced by the proto-type version of VPABase is the VPA-input as shown in Figure 10.

6 DATA-STORAGE AND DATA-EXCHANGE

Data are stored in tables and relations in the relational database, as described above. It is tradition in ICES to store data in so-called *flat-files*, that is a data-storage where each data record can be read independently, and where no relations between data are specified.

Data stored as flat-files usually takes up much more space than data stored as tables with relations. Should a flat-file be required, for example for analysis by the SAS, the database can easily produce such a flat-file.

Data can be submitted from countries to the central database in three ways.

1. Entry of data in replicate database
2. Entry of data in EXCEL format (Using a fixed standard set-up)
3. Entry of data from screen through the data entry forms

The data that is presented in Excel format, will be imported by the data base and added to the data already on the system. Alternatively, the database could have been implemented on the internet, and data could be entered directly in the central database. This solution is rather straight forward to implement.

The supply of data in a traditional “flat-file” format (e-g- SAS files) will be accommodated in the future.

The proto-types of the Excel data-exchange formats corresponds to the tables of VPABase: Country file (Table 12) and Fleet/age file (Table 13)

The fleet-table and the age-distribution table of VPABase have been combined into one exchange file. The fields in this table may not all be filled in. For example, if catch-at-age and weight-at-age are given, the landings (or discards) are given as the sum of products, and these cells do not have to be filled in. If they are, the SOP will have priority over the landings (discards) in tonnes.

This version of VPABase is not an “online-system”, and when a record is updated, it is not immediately updated in the international network. Each institute of participating in the ICES assessment working group will get a replica of the VPABase. A replica is simply a copy of VPABase, for which the design cannot be modified. Before the annual WG meeting, the master VPABase will be synchronised with all the replicas. To “synchronise” two replicas of a database, means to exchange all new records and updates of records. This task could be carried out by a data coordinator.

7 DISCUSSION AND CONCLUSIONS

A general conclusion from the whole project EMAS (Pastoors et al., 2001), was that the sampling procedures in the different countries have very different characteristics, adapted to the local circumstances, the availability of sampling options and the behaviour of the fishery. There is no obvious method for standardization of sampling procedures available. The implications of this uncertainty on stock assessment results are relatively minor when considering only the total catch at age matrix, but are larger when using commercial CPUE to tune the assessment model.

The storage of all basic data used in the estimation of catch at age is of great importance to keep track of the procedures used and record the consistency of sampling data. This is not yet the case in ICES, where different working groups use different methods and software, and where catch data are exchanged only at the national level of aggregation and with different format. The database presented here should therefore improve the homogeneity, standardisation, centralisation, transparency and flexibility of data used in assessment. It is generic and thus suitable for a great variety of fisheries. It allows accounting for various levels of aggregation and various scales of basic data. It gathers both the database and the program performing all calculations until the final VPA-input within one single software, with a great level of specification provided by the user. It has been designed with a great user-friendliness and easiness to use. It is not a rigid structure, as all user can have access to the programming code and, possibly, improve it for his own use.

However, VPABase is, at this stage of implementation, only a prototype, that requires both improvements and testing before being considered as possible standard software in ICES. The first improvement would be the ability to output not only the matrices of catches and weight at age, but also the tuning files. That would require extending the fleet record level to account for effort data, and also integrating survey data.

The second improvement would be to develop VPABase into a fully-integrated software that would perform the whole assessment procedure, instead of producing only input files to be used in existing software. That would avoid all problems due to format specifications and changes from one software to another, as it is currently the case in ICES working groups. The usual software used in fish stock assessment could be rewritten in a module in VPABase (e.g. in Visual Basic), which would become a full assessment package. Here again, the advantage of using ACCESS is the possibility for the user to see and modify the programming code, and thus to use the database and assessment programs with high flexibility and transparency.

The third improvement would be to add a plat-form independent module to the database, and to improve its access and use. Although ACCESS is widely used by a number of institutions, problems arise when using another operating system, or when updating the database (then a replica has to be sent to each individual partner each time). The best solution would be to provide an Internet access to VPABase, allowing using it from any kind of remote computer.

Finally, the database should be extensively tested by a number of working groups and fisheries research institutes before being validated as a standard tool. So far, the database has been tested only at DIFRES (Danish Institute for Fisheries Research). We believe that this tool could be of great help within ICES working groups, and we hope that it will encounter interest enough to be tested and used.

8 REFERENCES

- ICES, 2000 a). Report of the working group on the assessment of mackerel, horse mackerel, sardine and anchovy. ICES C.M. 2000 / ACFM:5.
- ICES 2000 b). Report of the Herring Assessment Working Group for the Area South of 60°N. ICES CM 2000/ACFM:10.
- Lewy, P., Lassen, H., 1997. Should total landings be used to correct estimated catch in numbers or mean weight-at-age? ICES J. Mar. Sci., 54: 462-470.
- Pastors, M.A., O'Brien, C.M., Flatman, S., Darby, C.D., Maxwell, D., Simmonds, E.J., Degel, H., Vinther, M., Sparre, P., Vanderperren, E., 2001. Evaluation of **MA**rket **S**ampling strategies for a number of commercially exploited stocks in the North Sea and development of procedures for consistent data storage and retrieval (EMAS). Final Report of the EU study no. 98/075. RIVO (The Netherlands), CEFAS (UK), DFU (Denmark), SOAEFD (Scotland), CLO-DZ (Belgium).
- Patterson, K. R. (1998). A Programme For Calculating Total International Catch-at-age and Weight at Age.

a	Age group
$a_{\text{Max}}(\text{St})$	Oldest age group of stock St , which may be a plus group
$a_{\text{Min}}(\text{St})$	Youngest age group of stock St (≥ 0)
Ar	Area index, $\text{Ar} = 1, 2, \dots, \text{Ar}_{\text{Max}}$
Ar_{Max}	Ar_{Max} = Number of areas.
$\text{Catch}(\Omega, \bullet)$	The total number caught of all age groups in stratum Ω .
$\text{Catch}(\Omega, a)$	Catch (Landings + discards) in numbers of age a in stratum Ω
Ct	Country
$\text{CV}_C(\Omega, a)$	Coefficient of variation of the estimate of $\text{Catch}(\Omega, a)$
$\text{CV}_{\text{wC}}(\Omega, a)$	Coefficient of variation of the estimate of $\text{W}_C(\Omega, a)$
$\text{C}_{\text{VPA}}(\text{St}, y, q, a)$	Input to VPA (Working group data)
$\text{DiscardN}(\Omega, a)$	Numbers discarded in stratum Ω , age group a (Working group data)
$\text{DiscardW}(\Omega, a)$	Yield discarded in stratum Ω (Working group data)
Fl	Fleet index, $\text{Fl} = 1, 2, \dots, \text{Fl}_{\text{Max}}$
$\text{Fl}_{\text{Max}}(\text{Ct})$	Fl_{Max} = Number of fleets of country Ct
$\text{LandingsN}(\Omega, a)$	Numbers landed in stratum Ω , age group a (Working group data)
$\text{LandingsW}(\Omega, a)$	Yield landed in stratum Ω , age group a (Working group data)
$\text{LandingsW}_{\text{Official}}(\text{St}, \text{Ct}, \text{Ar})$	Weight of annual landings by country Ct . Official statistics delivered from governments to ICES.
m_s	Number of sampled strata
M_t	Method of weighting sampled strata to estimate unsampled distributions
m_U	Number of un-sampled strata
$n_a(\Omega, a)$	Number of age measurements (otoliths)
$n_l(\Omega, a)$	Number of length measurements
$\text{Pr}_{\text{Catch}}(\Omega, a)$	The proportion made of age group in the total number
$\text{Pr}_{\text{Catch}}^{\text{Mean}}(\Omega_1^U, a, S_1, M_t)$	Weighted mean, using the set of samples S , and the weighting method M_t :
q	Quarter (or index of any other division of the year)
S	$S = \{\Omega_1, \Omega_2, \dots, \Omega_m\}$, set of m strata.
S^{All}	Set of all samples. $S^{\text{All}} = S^{\text{sampled}} \cup S^{\text{un-sampled}}$
S^{Sampled}	Set of sampled strata.
$S^{\text{un-Sampled}}$	Set of un-sampled strata.
$\text{SOP}_{\text{Catch}}(\Omega)$	Computed weight of the catch in a stratum Ω (Sum Of Products)
St	Stock (a sub-set of a species)
$\text{SW}_{M_t}(S_k)$	Set of weights associated with the set of sampled strata, selected by method M_t . $\text{SW}_{M_t}(S_k) = \{v_{M_t}(\Omega_1^S), v_{M_t}(\Omega_2^S), \dots, v_{M_t}(\Omega_k^S)\}$
$\text{W}_C(\Omega, a)$	Mean body weight in Catch of age group a in stratum Ω .
$w_C^{\text{Mean}}(\Omega)$	The mean weight of all fish in the catch in stratum
y	Year
$\text{Yield}(\Omega)$	Yield, Catch in weight in stratum Ω , (Working group data)
Ω	Stratum: $\Omega = (\text{St}, \text{Ct}, \text{Fl}, \text{Ar}, y, q)$. combination of Stock St , Country Ct , Fleet Fl , Area Ar , year y and quarter q .
Ω^S	Sampled stratum .
Ω^U	Un-sampled stratum.
$v_{M_t}(\Omega, a)$	Weighting factor for weighted mean of age group a , using method M_t

10 TABLES

Table 1. List of options for weighting the calculation of mean proportions at age and mean weight in the catch.

M t	Method	V_{Mt}(Ω,a)	Comment
1	Equal weight	1.0	No weighting
2	Catch	Catch(Ω, a)	Numbers caught (Cannot be used if “own” sample is used)
3	Landings	LandingsN(Ω, a)	Numbers landed
4	Discards	DiscardsN(Ω, a)	Numbers discarded
5	Yield (Landings or discards)	Yield(Ω) (LandingsW(Ω,) or DiscardsW(Ω,))	Weight of catch (Landings or discards)
6	Number of otoliths	n _c (Ω, a)	Problematic (see below)
7	Inverse CV (Age)	1/CV _c (Ω, a)	Only for catch-at-age (e.g. from Bootstrap)
8	Inverse CV(weight)	1/CV _{wc} (Ω, a)	Only for weight-at-age (e.g. from Bootstrap)

Table 2. Fields in input-data-tables of VPABase.

COUNTRY CATCH RECORD (Name in ACCESS: T_Country_Catch , see Fig. 3)	
Name of Field	Comment
<i>Key_Country_Catch</i>	Unique key of country record
Country	Look-up, name of country
Species	Look-up (A species can be separated into "Stocks")
Year	..., 1999,2000,2001,....
ICES Sub division	Look-up, name of ICES subdivision
ICES Division	Look-up, Not filled in if above filled in
ICES Area	Look-up, Not filled in if above filled in
Official Landings	As reported to ICES by governments
Unreported landings	WG members estimate of unreported landings
Misallocated landings	WG members estimate of landings
Discards	WG members estimate of discards
Sampled Catch	The weight of the catch covered by samples

FLEET CATCH RECORD (Name in ACCESS: **T_Fleet_Catch**, see Fig. 3)

Name of Field	Comment
<i>Key_Catch</i>	Unique key of fleet record. Also Pointer to sample (if existing) which was taken from this catch.
<i>Key_Country_Catch</i>	Pointer to country record (one-to-many)
Region (of country)	Part of country where home ports of fleets are (optionally)
Month	Look-up
Quarter	Look-up, Not filled in if month filled in
Fleet	Look-up, Sub set of fishing harvesting sector of country
Stock	Look-up, Subset of "Species"
Use of Catch	Look-up, Discards/Landings
Date or recording	The date the record was entered (or modified) in the database
Weight of Catch	Weight of landings or discards (catch = landings + discards is not input)
Number of samples	Number of age composition samples (=0 if no samples available)
Method of weighing	Index for weighing method. Default value = 1 for equal weight for each sample, updated when new selection is made
Otoliths	Sum of number of otoliths, updated when new selection is made
length measurements	Sum of numbers of length measurements, updated when new selection is made

AGE COMPOSITION RECORD (SAMPLE) (Name in ACCESS: **T_Age_Distribution**, see Fig. 3)

Name of Field	Comment
<i>Key_Age_Distribution</i>	Unique key of age sample record
<i>Key_Catch</i>	Pointer to fleet record (one-to-many)
Age group	0,1,....
Coefficient of variation	Variation of estimate of this age group (e.g. From bootstrapping)
Number of otoliths	Number of otoliths in age composition sample in this age group
Total number in catch	In discards or landings (raised numbers)
Mean weight	Mean weight of age group during time period in question

POTENTIAL USE OF SAMPLES FOR ESTIMATION OF AGE COMPOSITION *Auxiliary data-table, to contain information on the potential selection of samples to represent fleet records (catches). Contains all combinations of fleet-record and Age-composition-Record, whether selected or not.*

Name in ACCESS: **T_Use_of_Samles**,(see Fig. 3)

Name of Field	Comment
<i>Key_Catch_Age</i>	Pointer to age composition record (sample)
<i>Key_Catch</i>	Pointer to fleet record
Sample selected	Yes/No Yes: If a sample is selected to represent a catch, that is if the record “Key_Catch_Age” is selected to represent the fleet record “Key_Catch”
No sample	Yes/No. No: If no sample (age composition record) has been selected for the fleet Record “Key_Catch”

ACTUAL USE OF SAMPLES FOR ESTIMATION OF AGE COMPOSITION

Data-table, containing information on the actual selection of samples to represent fleet records (catches). Contains only combinations of fleet-record and Age-composition-Record, which are selected for estimation of age composition.

Name in ACCESS: **T_Use_of_Samles_Saved**,(see Fig. 3)

Name of Field	Comment
<i>Key_Catch_Age</i>	Pointer to age composition record (sample)
<i>Key_Catch</i>	Pointer to fleet record
Sample selected	Yes/No Yes: If a sample is selected to represent a catch, that is if the record “Key_Catch_Age” is selected to represent the fleet record “Key_Catch”
Weighting method	Look-up. Method used to compute weighted average of two or more samples.
Time for selection	The time and date the selection was made (in the database)

Table 3. Inventory of Country records (Example of report in ACCESS-format).

Inventory of Country Records

Country: Denmark								
Division:	Sub-Division:	Official landning	Mis-Allocated:	Un-reported:	Discards:	No. of fleets:	No of sampled fleets	Procent of catch sampled
IVa	IVa East	3379	0	0	0	4	0	100%
IVa	IVa East	14154	3506	0	0	4	1	166%
IVa	IVa Wes	652	0	0	0	4	0	100%
IVa	IVa Wes	22152	3506	0	0	4	0	116%
IVb	All	3906	0	0	0	4	0	30%
IVc	All	10648	3506	0	0	4	0	133%

Country: Netherlands								
Division:	Sub-Division:	Official landning	Mis-Allocated:	Un-reported:	Discards:	No. of fleets:	No of sampled fleets	Procent of catch sampled
IVa	IVa Wes	1000	100	100	200	1	1	1400%

Table 4. Inventory of fleet records

		LANDINGS			DISCARDS		
		Total yield	Number of otoliths	Coeff.of Vari-ance	Total discards	Number of otoliths	Coeff.of Vari-ance
Country A	Fleet 1	Yes/No	n	Yes/No	Yes/No	n	Yes/No
	Fleet 2	Yes/No	n	Yes/No	Yes/No	n	Yes/No

	Fleet x	Yes/No	n	Yes/No	Yes/No	n	Yes/No
Country B	Fleet 1	Yes/No	n	Yes/No	Yes/No	n	Yes/No
	Fleet 2	Yes/No	n	Yes/No	Yes/No	n	Yes/No

	Fleet x	Yes/No	n	Yes/No	Yes/No	n	Yes/No
.....							
Country X	Fleet 1	Yes/No	n	Yes/No	Yes/No	n	Yes/No
	Fleet 2	Yes/No	n	Yes/No	Yes/No	n	Yes/No

	Fleet x	Yes/No	n	Yes/No	Yes/No	n	Yes/No

Table 5. Report (proto-type): Country records. Official and non-official catch data.

Country records of ICES Division No. X					
	Official Landings	Unreported	Mis-allocated	Discards	Total
Country 1					
Country 2					
....
Country X					
Total					

Table 6. Report (proto-type): Fleet record (sample). Catch-at-age, weight-at-age and additional information on sampling (number of otoliths and coefficients of variation).

Country A Fleet a ICES Division No. x (sample No. y)									
Age gr.	LANDINGS				DISCARDS				TOTAL Number caught
	Number landed	Body wgt Landings	Number of otoliths	Coefficient of variation	Number Discards	Body wgt Discards	Number of otoliths	Coefficient of variation	
0									
1									
...
Max a									
TOTAL	SOP	Mean Wgt	Total	-	SOP	Mean Wgt	Total	-	SOP
If data above are not available, then total catc, landings or discards (in weight) may be available:									
Yield		Not avail.	Not avail.	Not avail.		Not avail.	Not avail.	Not avail.	

Table 7. Report (proto-type): Processing of fleets records or country records with incomplete samples.

Method of weighting						Fleet record No. X		Fleet record No. Y	
						Method no.	Method no.	Method no.	Method no.
Sample	Ct	Fleet	ICES-Div.	Land. Samp.	Disc. Samp.	Landings Samp. Used	Discards Sampl. Used	Landings Samp. Used	Discard Sampl. Used
1				Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No
2				Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No
.....
Z				Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No

Table 8. Report (proto-type): Results of processing of fleet-records and country-records.

Country A Fleet a ICES Division No. x						
Age gr.	LANDINGS			DISCARDS		TOTAL Number caught
	Number landed	Body wgt Landings		Number Discards	Body wgt Discards	
0						
1						
...
Max a						
TOTAL	SOP		Mean Wgt	SOP	Mean Wgt	SOP
Estimation	Method no.		Method no.			
Sample	Landings Samp. Used		Discards Sampl. Used			
1	Yes		Yes			
2	Yes		No			
.....			
Z	Yes		Yes			

Table 9. Report (proto-type): Country total of fleet records.

Country A All Fleets All ICES Divisions							
Age gr.	LANDINGS			DISCARDS			TOTAL Number caught
	Number landed	Body wgt Landings	Number of otoliths	Number landed	Body wgt Discards	Number of otoliths	
0	Σ	Σ	Σ	Σ	Σ	Σ	Σ
1	Σ	Σ	Σ	Σ	Σ	Σ	Σ
...
Max a	Σ	Σ	Σ	Σ	Σ	Σ	Σ
TOTAL	SOP	Mean Wgt	-	SOP	Mean Wgt	Total	SOP

Table 10. Input table to stock assessment (last column).

Age gr.	Number Landed	Number Discarded	TOTAL INPUT TO VPA
0	Σ	Σ	Σ
1	Σ	Σ	Σ
...
Max a	Σ	Σ	Σ

Table 11. Proto-type of Country record for data exchange.

COUNTRY RECORD	
Country	Denmark
Area	IV (Given only if Div or sub-div not filled in)
Division	IVc (Given only if sub-div not filled in)
Sub-division	IVc East
Year	2001
Official Landings	9000
Unreported	7000
Misallocated	3000
Discards	5000

Table 12. Proto-type of fleet record for data exchange.

FLEET RECORD

Country	Denmark			
Fleet	Trawlers <i>(Can be all Fleets, or set of fleets)</i>			
Area	IV <i>(Given only if Div or sub-div not filled in)</i>			
Division	IVc <i>(Given only if sub-div not filled in)</i>			
Sub-division	IVc East			
Year	2001			
Time period	Q1			
Land. (tons)	29135 <i>(when SOP available, this field need not be filled in)</i>			
Disc. (tons)	38692 <i>(when SOP available, this field need not be filled in)</i>			
	Total number landed/ discarded	Number of Otoliths	Average Body Weight (during period)	Coefficient of Variation %
Land Age 1	564	6	1.10	12
Land Age 2	1128	12	1.50	5
Land Age 3	1034	11	2.00	8
Land Age 4	282	3	2.20	15
Discard Age 1	2883	12	0.95	16
Discard Age 2	512	8	1.15	25
Discard Age 3	0	0		
Discard Age 4	0	0		

Figure 1. Illustration of the gap-filling-process.

The “borrowing” of samples is illustrated in the figure by the broken lines from a sample to a catch-age-composition.

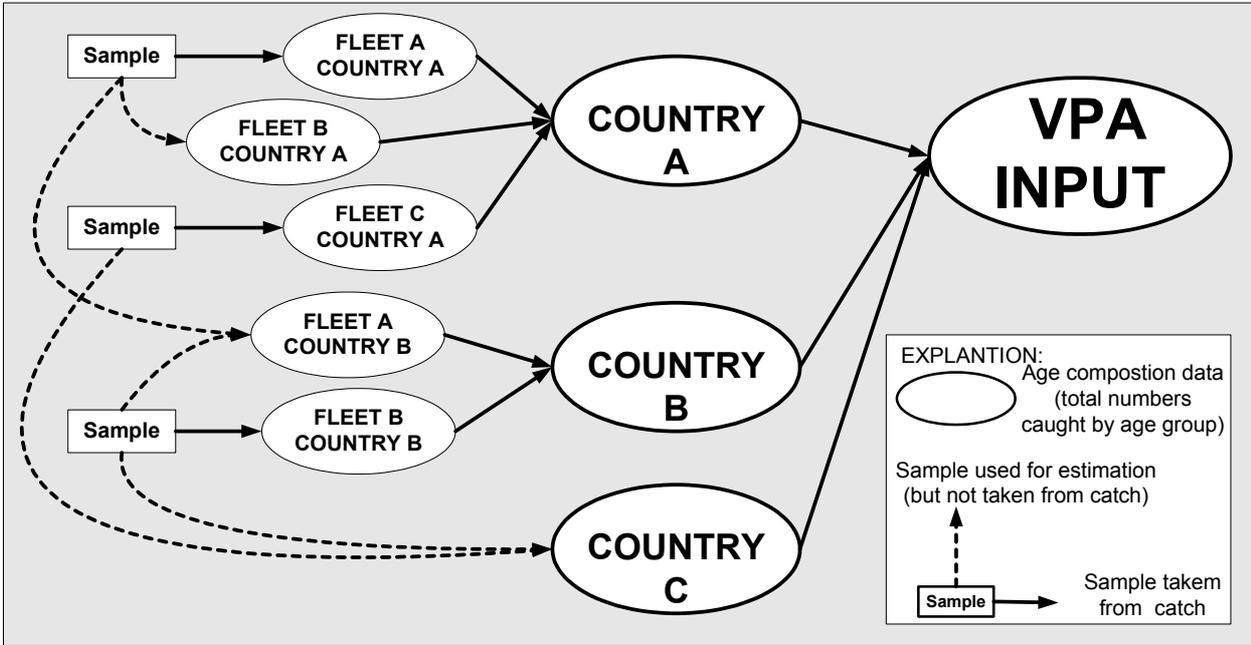


Figure 2. The basic hierarchy of tables in VPABase

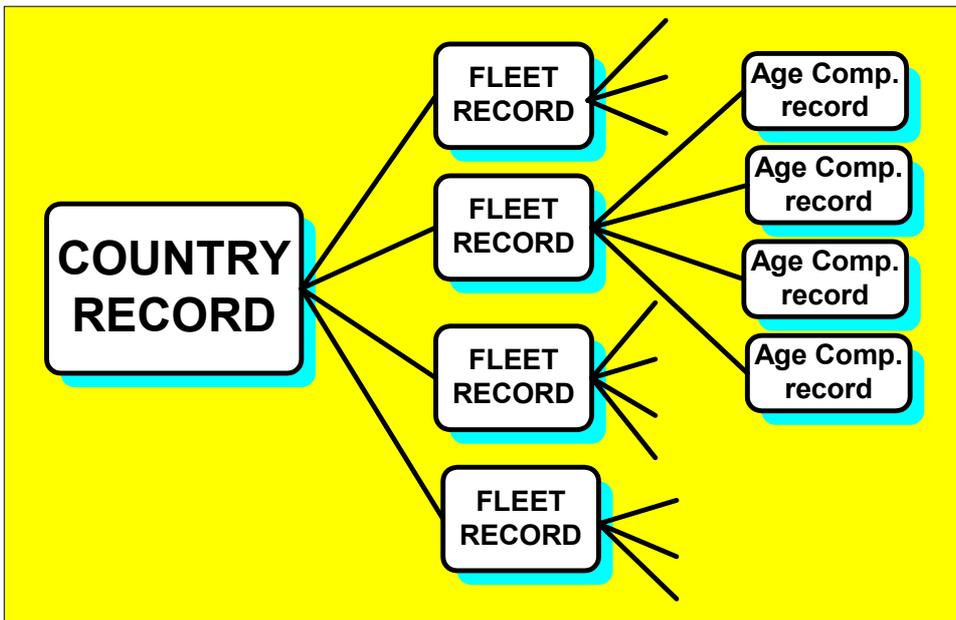
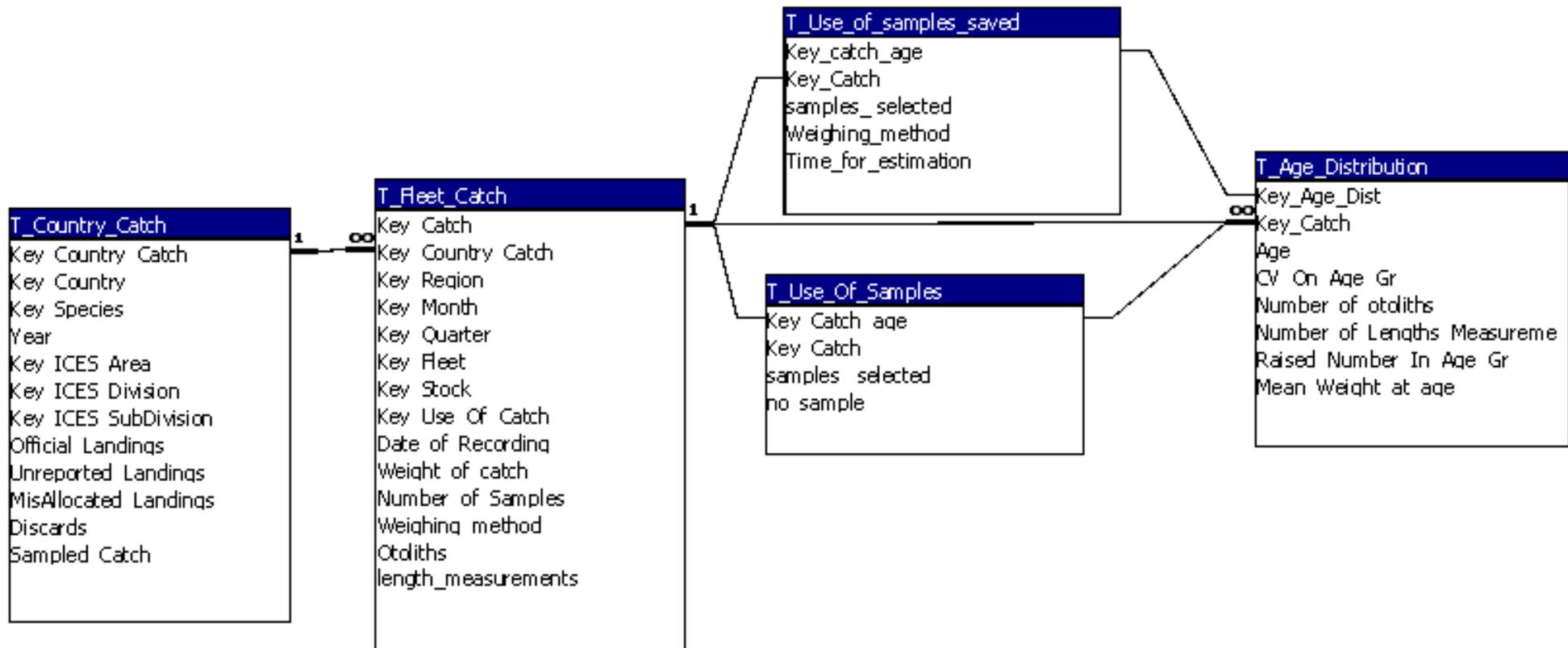


Figure 3. Data Tables and relations of the ACCESS implementation of VPABase

(Compare Table 1)

Note that the two tables “T_Use_Of_Samples” and “T_Use_Of_Samples_Saved”, are the tables containing information on the data-processing, whereas the other tables contain



“observations” from fishery.

Figure 5. Main menu of VPABase.

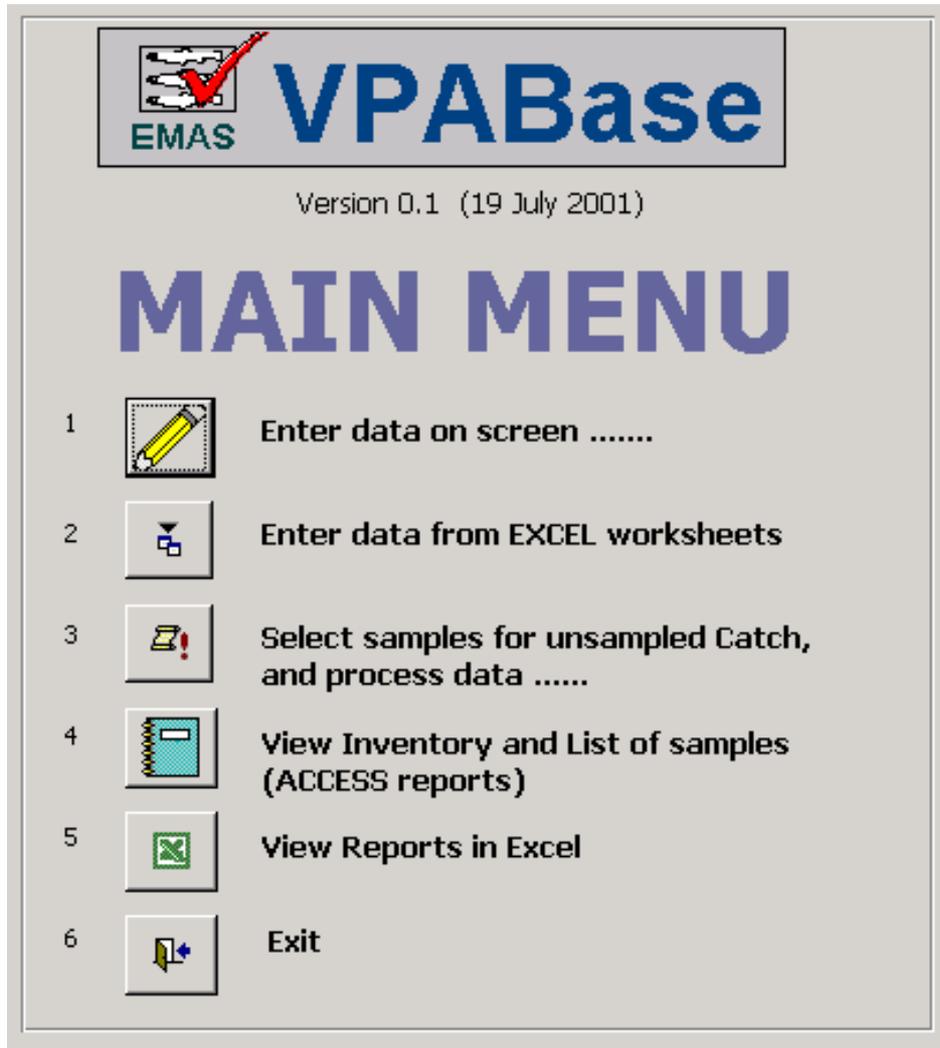


Figure 6. Data entry form of VPABase.

catch
_ □ ×

Data entry

VPABase

ANNUAL CATCH BY COUNTRY

Country	<input type="text" value="Netherlands"/>	Official Landings:	<input type="text" value="469"/>	(tons)
Year	<input type="text" value="2001"/>	Unreported Landings:	<input type="text" value="75"/>	
Species	<input type="text" value="Cod"/>	MisAllocated Landings:	<input type="text" value="-82"/>	
ICES Sub Area:	<input type="text" value="IV"/>	Discards:	<input type="text" value="15"/>	
ICES Division:	<input type="text" value="IVc"/>	Sampled Catch:	<input type="text" value="200"/>	
ICES Sub Division:	<input type="text" value="All"/>			

Save and Exit

Add new record

CATCH BY FLEET AND TIME PERIOD

Quarter	<input type="text" value="Q1 (Jan-Feb-Mar)"/>	Use Of Catch	<input type="text" value="Reduction"/>
Fleet	<input type="text" value="OB Trawl"/>	Number of samples:	<input type="text" value="5"/>
Stock	<input type="text" value="North Sea Cod"/>	Weight of Catch:	<input type="text" value="61"/>
		Date of Recording:	<input type="text" value="28-jun-01"/>

Add new record

AGE COMPOSITION SAMPLE

W-rings	Catch No.(000)	Body Weight (kg)	No. Lengths M.	No otoliths	CV (%)
<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
<input type="text" value="1"/>	<input type="text" value="8800"/>	<input type="text" value="0,015"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
<input type="text" value="2"/>	<input type="text" value="2930"/>	<input type="text" value="0,044"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
<input type="text" value="3"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
<input type="text" value="4"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
<input type="text" value="5"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>
<input type="text" value="6"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>

Record: ⏪ ⏩ ▶▶▶* of 4

Record: ⏪ ⏩ ▶▶▶* of 12

Figure 7. Menu for data processing

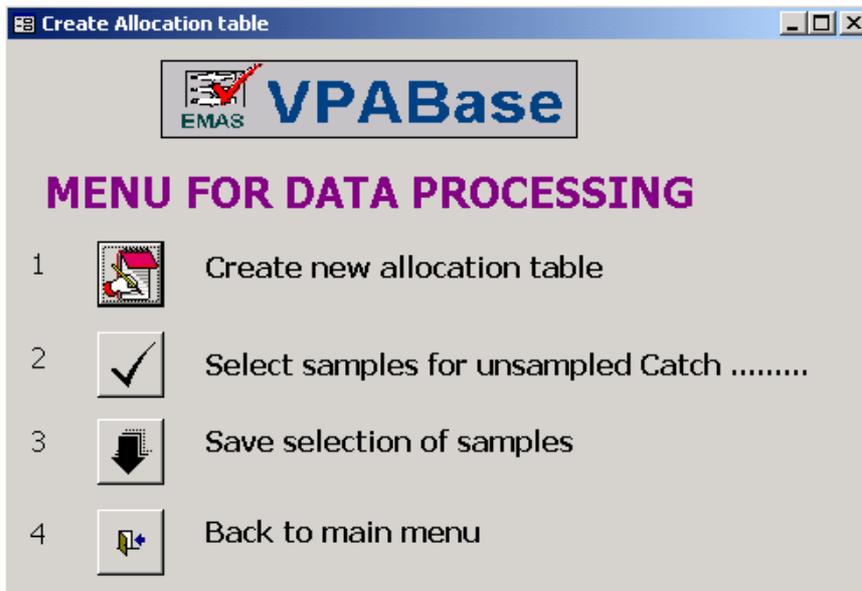
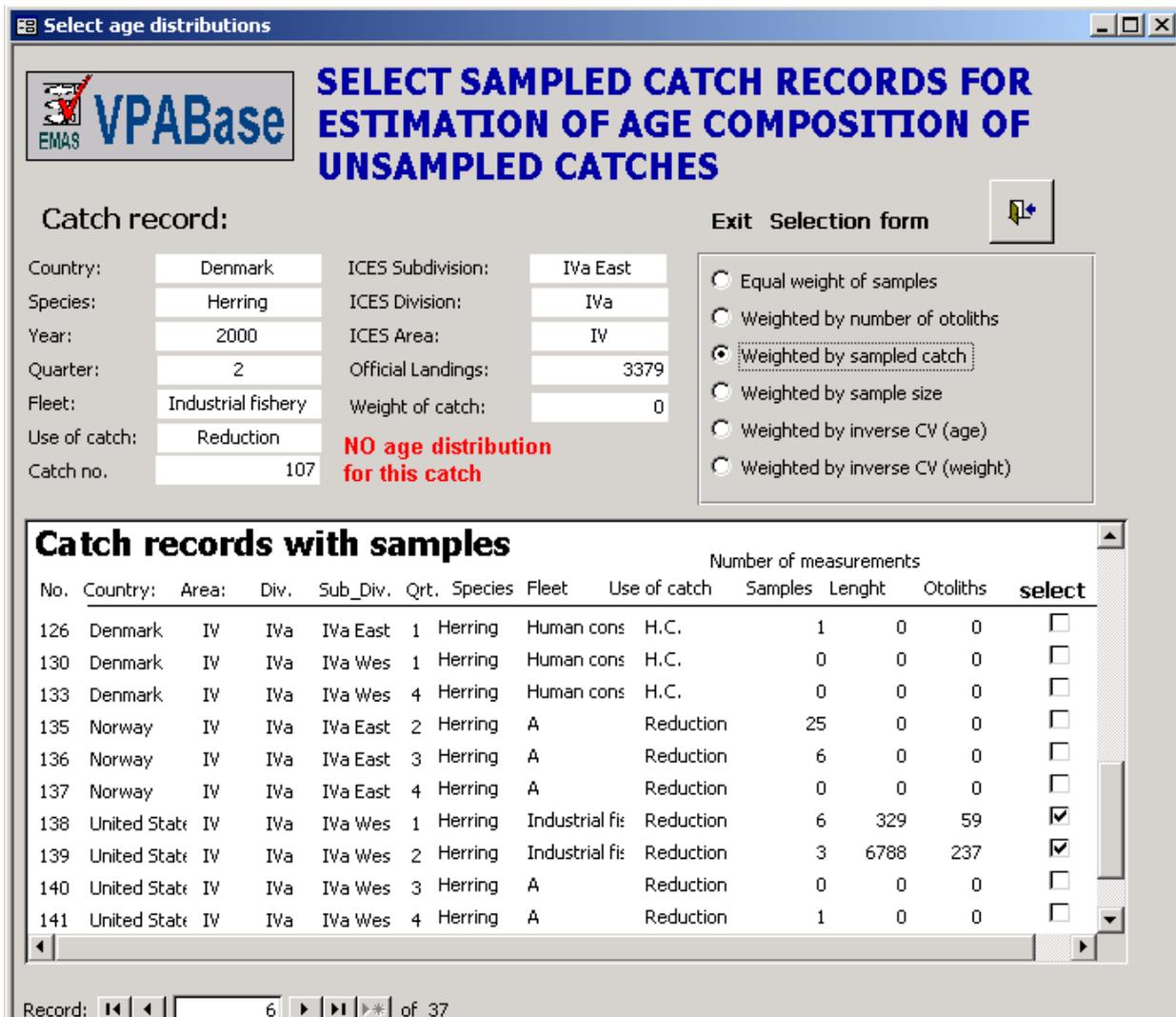


Figure 8. Data entry form for the selection of age compositions to be used to fill-in unsampled catches



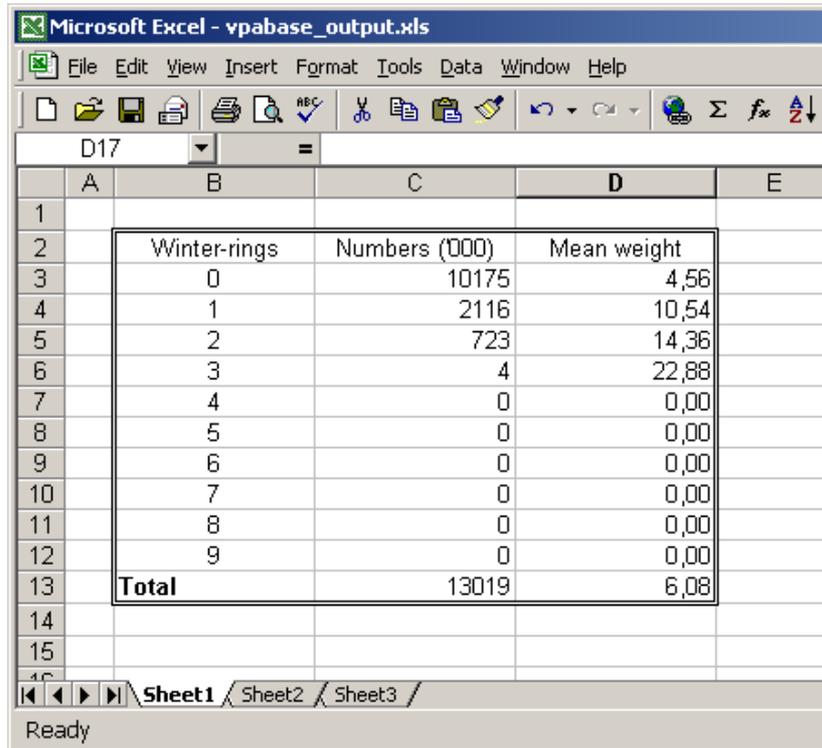
(hypothetical example).

Figure 9. EXCEL worksheet used by the ICES Herring WG and the Mackerel, Horse mackerel, Sardine and Anchovy WG.

(The figure shows only a part of the worksheet)

CATCH NUMBERS, LENGTH AND WEIGHT AT														Catch data (WG estimate) should include discards and unre	
WG for the Assessment of Herring in the Area south of 62°														TONNES SHOULD ALWAYS BE FILLED IN (even if no numbe	
Country:	Denmark			Division:	IVaW										
Species:	Herring			Fleet:	Industrial fishery										
Year:	2000			Revised (date):	3-mar-'00			*only t year is							
		Quarter 1 2000			Quarter 2 2000			Quarter 3 2000			Quarter 4 2000			Qual	
Winter-Rings	Numbers at age ('000)	Mean Length (cm)	Mean Weight (kg)	Numbers at age ('000)	Mean Length (cm)	Mean Weight (kg)	Numbers at age ('000)	Mean Length (cm)	Mean Weight (kg)	Numbers at age ('000)	Mean Length (cm)	Mean Weight (kg)	Numbers at age ('000)	Mean Length (cm)	Mean Weight (kg)
0							37	11,5	0,014	651	11,5	0,014			
1							1607	21,5	0,064	28 171	21,5	0,064			
2							166	25,0	0,116	2 909	25,0	0,116			
3	159	26,5	0,114				238	26,2	0,133	4 172	26,2	0,133			
4	635	28,0	0,140				52	26,5	0,140	919	26,5	0,140			
5							41	25,5	0,112	727	25,5	0,112			
6															
7															
8	159	30,5	0,190				41	31,5	0,206	727	31,5	0,206			
Plus-group	9+														
Total/Mean -->	954	28,2	0,144		na	na	2 182	22,5	0,080	38 276	22,5	0,080			
Catch	137	(t)		Catch	na	(t)	Catch	175	(t)	Catch	3 067	(t)			
SoP	137	(t)		SoP	na	(t)	SoP	175	(t)	SoP	3 064	(t)			
SoP	100%	(%)		SoP	na	(%)	SoP	100%	(%)	SoP	100%	(%)			
No. aged	2	(n)		No. aged	0	(n)	No. aged	0	(n)	No. aged	13	(n)			
New sheets for other fishing areas may be copied by <EDIT> <MOVE OR COPY SHEET>.															
Then rename the sheet according the new fishing area by double clicking on it. (Rename also this CANUM sheet at															
CANUM IVaE CON / CANUM IVaW CON / CANUM IVaE IND / CANUM IVaW IND / CANUM IVb IND															

Figure 10. Example of report in EXCEL format.



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - vpabase_output.xls". The menu bar includes File, Edit, View, Insert, Format, Tools, Data, Window, and Help. The toolbar contains various icons for file operations and editing. The active cell is D17. The spreadsheet displays a table with the following data:

	A	B	C	D	E
1					
2		Winter-rings	Numbers ('000)	Mean weight	
3		0	10175	4,56	
4		1	2116	10,54	
5		2	723	14,36	
6		3	4	22,88	
7		4	0	0,00	
8		5	0	0,00	
9		6	0	0,00	
10		7	0	0,00	
11		8	0	0,00	
12		9	0	0,00	
13		Total	13019	6,08	
14					
15					
16					

The status bar at the bottom shows "Ready" and the sheet tabs "Sheet1", "Sheet2", and "Sheet3".

APPENDIX 6



Working Document: PGCCDBS/WD: XX

ICES Planning Group on Commercial Catch, Discards and Biological Sampling [PGCCDBS]
Lisbon, 5-8 February 2002

Some ideas to coordinating and evaluating EU sampling efforts

Martin Pastoors

Netherlands Institute for Fisheries Research (RIVO-DLO)

P.O. Box 68, 1970 AB IJmuiden, The Netherlands

+ 31 255 564690 / + 31 255 564644 (fax)

m.a.pastoors@rivo.wag-ur.nl

Introduction

The Planning Group on Commercial Catch, Discards and Biological Sampling [PGCCDBS] was set up at the Annual Science Conference 2001 with the aim to provide a forum for coordination of the range of sampling activities within the Western European waters. The direct incentive to set up the planning group was the new EU regulation on data collection (EC 2000) which requires each country to sample landings (and discards) of not only the vessels that fly the flag of that country, but also to sample from 'foreign' vessels that land into that country. The planning group will be instrumental to provide coordination of these sampling activities and also to provide guidelines on how the resulting age compositions can be transferred to the relevant assessment working groups.

The specific terms of reference set for this first meeting of the planning group are as follows:

- a) evaluate the commercial catch (landings), discard and biological sampling programmes being implemented in the Baltic Sea, North Sea, Western and Southern waters, and assess whether or how these require ICES coordination;
- b) evaluate the need for developing sampling methodology, calculation methodology, data storage procedures, and software for aggregating national catch at age data to international catch at age in a form suitable for assessment working groups, and prepare relevant proposals and work plans.

In this working document I will outline some ideas to deal with the issues that are contained within the terms of reference. I will first describe the type of data and software needed to address ToR a). I will then go on to describe how we foresee the longer term trends in dealing with sampling standardization and the shorter term solutions to data exchange and calculation methodologies applicable to assessment working groups.

Data needed to evaluate sampling programmes

What we need is data for two purposes:

- to evaluate the sampling coverage for assessment purposes
- to evaluate the sampling in relation to the EU regulations.

These two purposes require data on different levels of spatial aggregation. To evaluate sampling for assessment purposes we would need the level of sampling by ICES subdivision and possibly also the type of sampling (e.g. age sampling, length sampling, age/length sampling). For the EU data regulation the main requirement is the number of age and length measurements by TAC area (usually a combination of subdivisions).

In order to achieve these objectives, I propose to make the following information available on a routine basis to the planning group:

- sampling information in the recent past (e.g. last three years)
- sampling information planned for 2002
- landings and discard data in the recent past
- TACs for 2002
- minimum sampling requirements according to the EU regulation

But how to store the data. In principle it would be possible to store data at the sample level. This would allow any type of aggregation to be made afterwards. However, this would also entail a fair amount of database work, which may not be needed. Therefore I propose to aggregate sampling and catch information at the ICES subdivision level

and by quarter, similar to the implementation in ISLDB. However we do need the additional data listed above. An example of the type of exchange format that would be needed is given in table 1.

One could wonder whether the same format could also be used for planning purposes, i.e. to assist in planning market and discard sampling within the planning group.

Exchange of working group data

The procedures for the preparation of data for assessment purposes will need to be changed in the light of the new sampling requirements. Notably the necessity to sample foreign vessels will mean that samples will be collected that apply to fleets for which no comprehensive data is available to the country of sampling. The way sampling is carried out by different countries is not standardized. Therefore I see no gain in exchanging sampling information to the country of flag, since the stratification of sampling may be specific to the country of sampling. I would like to suggest that the country of sampling also becomes the country of raising the sampling data to a standardized level. They would need to generate standardized age-length distributions (e.g. referring to one ton of landings). The country of flag would raise these to the appropriate fleet landings.

The software that is currently in use for the exchange and combination of national or fleet-based age compositions is often outdated and not equipped for the situation of multiple fleets and multiple countries. Several promising new approaches are now available (e.g. salloc, VPABase), and may be adapted to give the resolution needed.

Salloc is a Fortran program written by Ken Patterson (Patterson 1998). It has been applied since 1999 by the Mackerel WG (WGMHSA) and the Herring Assessment WG (HAWG). The program is intended to do the following:

1. Read data in a standardized format
2. Allow the data coordinators sufficient freedom in defining re-allocations of unsampled catches to specified age-structures
3. Perform standard calculations automatically and with no calculations done by the user.
4. Generate standard tables of catches by country and sampling intensity.
5. Document the choices made by the data coordinator in allocating age-structures to unsampled catches
6. Calculate the total international catch-at-age and weights at age.

Because treatment of stocks in respect of fleet units, areas and to time-periods can be rather different, the program allows flexible use of time, fleet and area units. The difficult part is to get the allocations right for the catches without samples. The Salloc program is very generic and runs under all windows operating systems.

The VPABase is an Microsoft Access 2000 implementation of a similar idea as Salloc (Pastoors et al. 2001; Sparre et al. 2001). In this case, Access is used as the database and the routines are programmed in Visual Basic. It reads the input directly from the Excel exchange formats developed by the Herring Assessment WG. The main draw-back from this approach is the reliance on Microsoft Access 2000 and the fact that the basic data files have a very complicated structure. However, it does provide a promising direction for further development.

Both for Salloc and VPABase, a major drawback is the format of the exchange files which is very complicated. An exchange format which can be delivered in a simple text format would be preferred.

Standardization of sampling procedures

We are now faced with a complicated situation where multiple countries take samples from multiple fleets. Ideally, it should be possible to exchange samples from one country with those of another country, so that the optimal use of the sampling effort could be made. This will only be feasible if the sampling protocols are designed in a generic fashion so that a sample taken by country X is the same type of sample as that by country Y. Otherwise the stratification will always limit the possibility of exchange of raw sampling data. In principle, I feel that the development of these standardized sampling protocols should be investigated in the near future. Perhaps this is only feasibly when using simple two stage sampling with relatively little stratifications (e.g. by ICES subdivision and quarter?). Theoretical studies are already available on the amount of sampling that would be needed to obtain an optimal allocation of sampling effort (e.g. Gudmundsdottir et al. 1988; Nicolajsen and Grastein 1993; ICES 1994; Reinert and Lewy 1999; Oeberst 2000)

Conclusions

1. In order to allow for the evaluation of sampling levels in the (recent) past and for the planning of sampling in the coming year(s) it is necessary to further develop and international database system that will contain meta-data on sampling in the different countries. A system like ISLDB could be used but it would need to be changed to accommodate data on TAC's and on the sampling requirements by the EU.
2. Sampling countries should generate age compositions by a certain standardized catch unit (e.g. ton), which would then be raised by the country of flag.
3. Exchange format for raising programs should be simple text format
4. Raising programs should be written in a language which is readily accessible to the majority of the users.
5. Standardization of sampling procedures that would allow the exchange of raw sample data, entails a route to lower levels of stratificatio and a uniform two stage sampling approach.

References

- EC (2000). COUNCIL REGULATION (EC) No 1543/2000 of 29 June 2000 establishing a Community framework for the collection and management of the data needed to conduct the common fisheries policy. **No. 1543/2000.**
- Gudmundsdottir, A. B., A. Steinarson and G. Stefansson (1988). A simulation procedure to evaluate the efficiency of some otolith and length sampling schemes. ICES C.M. 1988 / D:14.
- ICES (1994). Report of the workshop on sampling strategies for age and maturity, Copenhagen, 3-9 february 1994. **ICES C.M. 1994 / D:1 ref: G,H,J.**
- Nicolajsen, A. and J. M. Grastein (1993). What is the most cost-effective procedure for sampling landings from a commercial fishery? ICES C.M. 1993 / D:41.
- Oeberst, R. (2000). "An universal cost function for the optimization of the numbers of age readings and length measurements for age-length-key-tables (ALKT)." Archive of Fishery and Marine Research **48**(1): 43-60.
- Pastoor, M. A., C. M. O'Brien, S. Flatman, C. D. Darby, D. Maxwell, E. J. Simmonds, H. Degel, M. Vinther, P. Sparre and E. Vanderperren (2001). Evaluation of market sampling strategies for a number of commercially exploited stocks in the North Sea and development of procedures for consistent data storage and retrieval (EMAS). Final report, RIVO, CEFAS, DIFRES, SOAEFD, CLO-DZ. **CFP 98/075.**
- Patterson, K. R. (1998). A Programme For Calculating Total International Catch-at-age and Weight at Age.
- Reinert, R. and P. Lewy (1999). Optimum biological sampling in the Danish landings of human consumption species in the North Sea, DIFRES.
- Sparre, P., O. Folmer and C. M. Ulrich (2001). Prototype of database for storage and processing of VPA-input data. ICES C.M. 2001 / P:.

Table 1. Example of a exchange format for meta-data on catches and sampling.

Flag_country	Sampling_country	Landing_country	year	quarter	division	HARBOUR	GEAR	Catch_type	catch_ton	SPECIES	samples_n	meas_n	aged_n	sweight_kg	status
BC	BC	BC	1998	01	VIIIc	ALL	Baka Trawl	LAND	3.1	ANK	0	0	0	0	real
BC	BC	BC	1998	01	VII	ALL	Baka Trawl	LAND	0	ANK	3	226	0	212.1	real
BC	BC	BC	1998	01	VIIIa	ALL	Baka Trawl	LAND	29.6	ANK	4	404	0	238.3	real
BC	BC	BC	1998	01	VIIIa	ALL	Bou Trawl	LAND	5.5	ANK	0	0	0	0	real
BC	BC	BC	1998	01	VIIIa	ALL	Hooks	LAND	0	ANK	0	0	0	0	real
BC	BC	BC	1998	01	VIIIc	ALL	Pair Trawl	LAND	0.5	ANK	0	0	0	0	real
BC	BC	BC	1998	01	VIIIa	ALL	Pair Trawl	LAND	5.9	ANK	2	51	0	55.5	real
BC	BC	BC	1998	01	VIIIa	ALL	GillNets	LAND	0.1	ANK	0	0	0	0	real
BC	BC	BC	1998	01	VIIIa	ALL	Small GillNets	LAND	0.6	ANK	0	0	0	0	real
BC	BC	BC	1998	01	VIIIc	ALL	Small GillNets	LAND	0	ANK	0	0	0	0	real
BC	BC	BC	1998	01	VIIIc	ALL	Baka Trawl	LAND	7.1	HKE	0	0	0	0	real
BC	BC	BC	1998	01	VI	ALL	Baka Trawl	LAND	14.3	HKE	1	92	0	122.8	real
BC	BC	BC	1998	01	VIIIa	ALL	Baka Trawl	LAND	26.8	HKE	4	368	0	72.9	real
BC	BC	BC	1998	01	VIIIa	ALL	Bou Trawl	LAND	8.4	HKE	0	0	0	0	real
BC	BC	BC	1998	01	VIIIa	ALL	Purse Seine	LAND	0	HKE	0	0	0	0	real
BC	BC	BC	1998	01	VIIIc	ALL	Purse Seine	LAND	0	HKE	0	0	0	0	real
BC	BC	BC	1998	01	VIIIa	ALL	Hooks	LAND	0	HKE	0	0	0	0	real
BC	BC	BC	1998	01	VIIIc	ALL	Hooks	LAND	0	HKE	0	0	0	0	real
BC	BC	BC	1998	01	VIIIa	ALL	LongLiner	LAND	0	HKE	0	0	0	0	real
BC	BC	BC	1998	01	VIIIc	ALL	Pair Trawl	LAND	7.6	HKE	0	0	0	0	real
BC	BC	BC	1998	01	VIIIa	ALL	Pair Trawl	LAND	278	HKE	4	952	0	723.3	real
BC	BC	BC	1998	01	VIIIa	ALL	GillNets	LAND	30.1	HKE	2	221	0	371.1	real
BC	BC	BC	1998	01	VIIIa	ALL	Small GillNets	LAND	13.2	HKE	0	0	0	0	real
BC	BC	BC	1998	01	VIIIc	ALL	Small GillNets	LAND	4.8	HKE	0	0	0	0	real
BC	BC	BC	1998	01	VIIIc	ALL	Baka Trawl	LAND	30.7	HMM	0	0	0	0	real
BC	BC	BC	1998	01	VIIIa	ALL	Baka Trawl	LAND	7.8	HMM	0	0	0	0	real
BC	BC	BC	1998	01	VIIIa	ALL	Purse Seine	LAND	17.9	HMM	3	229	0	47.1	real
BC	BC	BC	1998	01	VIIIc	ALL	Purse Seine	LAND	87.4	HMM	2	81	0	23.7	real
BC	BC	BC	1998	01	VIIIa	ALL	Hooks	LAND	0	HMM	0	0	0	0	real

APPENDIX 7

Working Document for the PGCCDBS meeting in Lisbon, 5-8 February 2002

Proposed fields for new data base for monitoring age, length and discards sampling

Guus Eltink

RIVO, Netherlands

A very preliminary proposal for a data base for market sampling and discard sampling, which can be used for discussion during the meeting of APGCCDBS.

FIELD	DATA TYPE	COMMENT
Species ????	Text (xx char)	Latin/English name/3 char code
Sample_no	Integer (3 digits)	001 - 999
Sample_type	Text (10 char)	market, discards
Sample_weight_kg	Decimal (kg)	e.g. 22.3
Intitute_sampled	Text (xx char)	e.g. IFREMER
Numbers_measured	Integer (4 digits)	1 - 9999
Numbers_aged	Integer (3 digits)	1 - 999
Year	Integer (4 digits)	e.g. 2002
Month	Integer (2 digits)	01 - 12 = Month of <u>catch landed</u>
ICES_rectangle (if not achievable ICES_Div)	Text (5 char) (Text (xx char)	e.g. 36 F2 (e.g. 8b)
Country_sampling	Text (3 char)	3 char-codes e.g. NET
Country_flag	Text (3 char)	3 char-codes e.g. NET
Country_catch_landed	Text (3 char)	3 char-codes e.g. NET
Harbour	Text (xx char)	5 char-codes e.g. NET
Gear	Text (xx char)	3 char-codes ??
Vessel_segm_1 >=40m	Text (xx char)	<12 or 12-<24 or 24-<40 or
Vessel_segm_2 221kW	Text (xx char)	only for Beamtrawlers < or >=

Comment: Information by ICES rectangle would be beneficial if different kind of areas are used (e.g. ICES-Divisions versus Roundfish areas in North Sea or versus Herring Areas, etc. If the distribution of all stocks can be described by ICES-Divisions then ICES rectangles would not be necessary!

FIELD	DATA TYPE	COMMENT
Species ????	Text (xx char)	Latin/English name/3-char code
Stock	Text (xx char)	e.g. Anchovy Bay of Biscay
Stock_distribution_area	Text (xx char)	e.g. 7efghjk8abde
ICES_rectangle (if not achievable ICES_Div)	Text (5 char) (Text (xx char))	e.g. 20 E7 (e.g. 8b)
Minimum_level_AGE tonnes)	Decimal	e.g. 0.025 (=25 per 1000)
Minimum_level_LENGTH tonnes)	Decimal	e.g. 0.050 (=50 per 1000)

Comment: This set of data only needs to be updated if there are changes in the distribution of stocks and if the minimum EU-requirements for age- and length sampling have changed by stock.

FIELD	DATA TYPE	COMMENT
Species ????	Text (xx char)	Latin/English name/3-char code
Stock	Text (xx char)	e.g. Anchovy Bay of Biscay
Year	Integer (4 digits)	e.g. 2002
Month	Integer (2 digits)	01 - 12 = Month of <u>catch landed</u>
ICES_Division	Text (xx char)	e.g. 8b
Country_flag	Text (3 char)	3 char-codes e.g. NET
Country_catch_landed	Text (3 char)	3 char-codes e.g. NET
Vessel_seg1 >=40m	Text (xx char)	<12 or 12-<24 or 24-<40 or
Vessel_seg2 221kW	Text (xx char)	only for Beamtrawlers < or >=
Gear	Text (xx char)	3 char-codes ??
Landings_tonnes	Decimal	e.g. 2548 tonnes
Discards_tonnes	Decimal	e.g. 329 tonnes

Comment: This set of data needs to be updated if catch and discards data become available !

APPENDIX 8

Description of
DAtabase TRAwl Surveys

By Lena Inger Larsen

ICES Secretariat

1	Introduction	67
2	Database structure.....	68
3	Dataflow	69
4	Output and Internet access.....	70
5	Data exchange	71
6	Checking program	72
7	Calculating CPUE at age and length.....	72
7.1	ALK aggregation areas	73
8	Indices	73
8.1	Beam trawl survey	74
8.2	BITS Indices.....	74
8.3	IBTS indices	75
9	Species codes.....	75
10	Conclusion	76
	Appendix I: Standard tables in the BITS and IBTS database.....	83
	Appendix II: XML Validation	87
	Appendix III: Checks performed by the present checking program in the BITS and IBTS data.....	89
	Appendix IV: Calculations in the BITS database	93
	Appendix V: Calculations in the IBTS database.....	97

1 INTRODUCTION

This report is written for consideration by the Steering group of the DAtabase TRAwL Surveys project (DATRAS) at their first meeting in January 2002. The report is a draft on the DATRAS database design, the standard output it should provide and how calculations should be done. The report identifies a series of problems which should to be dealt with before the database can be developed.

The DATRAS database will contain data from the following surveys:

- The Baltic International Trawl Survey (BITS), Baltic Sea
- The International Bottom Trawl Survey (IBTS), North Sea, Skagerrak, Kattegat
- The International Bottom Trawl Survey (IBTS), western and southern divisions
- The Beam Trawl Survey (BTS), North Sea, Channel and Irish Sea

The database will be held in Microsoft SQL-server and access will be at two task levels. One level, which only will be accessible in ICES (database manager) where import, updates and maintenance are done, and one level for data extractions which is for all users and can be entered through the Internet.

The database should have the following functionalities:

- **Loading data**

The data will be sent to ICES from the national laboratories in a defined exchange format. At present, there are exchange formats for the IBTS North Sea, Skagerrak and Kattegat and for the BITS data. Exchange formats still need to be developed for BTS and IBTS western and southern divisions. The formats for IBTS and BITS are described in the IBTS and BITS manuals respectively.

Before sending the data to ICES they must be screened with a checking program by the national laboratories. At ICES the data will be checked once more before they are loaded into the database. A new checking program will be developed under the DATRAS program. For more information about the checking program see section 6.

- **Correcting/deleting data in the database**

Data in the present IBTS database are not updated in the database but in the exchange text files. Whenever data are needed from the database the latest version of the data are loaded and data are extracted. This is a very time demanding task and not how a database is supposed to work. In the BITS database there are created functionalities for updating data directly in the database. These are, however, not as well developed, as one could hope for.

In the DATRAS database these functionalities should be fully developed. All updates and corrections should be logged so the database administrator at any time can see when and where updates have been done. To ensure that data can be regenerated, the original data must still be kept in the exchange text files.

This approach has the same weakness as the earlier used approach in IBTS; if data are update in ICES there is no guarantee that the data submitters updates their data and if they later resubmit data earlier updates might have to be done again. The most safe and less time consuming approach on data update would therefore be that the data submitter are responsible for the quality of the exchange files and that they therefore would have to resubmit all data when mistakes are found in the data in ICES.

- **Producing standard indices**

One of the main tasks for the surveys is to provide indices used by various assessment working groups to tune the VPA. To standardise how the indices are calculated for the different species in the different surveys, the DATRAS database should be able to do these calculations. Section 8 gives an overview of the different indices that the database should be able to handle. How the different indices should be calculated for IBTS and BTS have to be specified at the first steering group meeting. The specifications for BITS can be found in Appendix IV.

- **Producing data output**

The database should be able to produce different levels of data output. The data output will be further dealt with in section 4 where the present data output from IBTS and BITS are described together with requested output from BTS. The future outputs from the DATRAS database have to be decided by the steering group at its first meeting.

The output from the database is closely related to the access level of the data. The access can be divided into four groups:

- Public
- Data suppliers
- Requests
- Database administrator (full access to the whole database, makes extractions on request)

The overall goal for the database is to provide data at the following levels:

- Public accessible aggregated data files. According to ICES data politics certain aggregated data are public available and in the future it should be possible to extract these data through the web.
- Non-aggregated data should be accessible through the web for the data suppliers. This means that laboratories delivering data from e.g. the Baltic Sea only have fully access to data from the Baltic Sea.
- Extracts on request. ICES hopes that a future web based database will reduce the number of requests for data considerably.
- Metadata for integration. ICES is working on creating a database system holding metadata from all databases in ICES. This system should allow everybody to search through the web and get information on what data are available in all databases in ICES. DATRAS should be able to provide the required metadata for such a system.
- Standard tables and graphs

2 DATABASE STRUCTURE

Data in all surveys are given haul-by-haul as mean catch per unit of effort (CPUE) per length class per hour. In addition, number at age, maturity and sex per length class are given for sub-samples of the target species (table 1, section 8). The sub-samples are given either haul-by-haul, ICES statistical rectangle or sampling area (e.g. roundfish area) depending on survey.

The database will be a relational database. The database design shown in figure 1 is designed from the data available in the present exchange format (IBTS and BITS). These data are, as noted above, calculated as mean per one hour trawling. At the IBTS meeting 2001 there was a request for additionally including sub samples of data per haul in the exchange format. If it is decided to include data at that level the table tblSubSample in figure 1 will have to be fully developed in the database design and dataflow would have to be extended to handle these new haul based data.

All data fit into a relational database structure except the area defined age/length key (ALK) in the IBTS and BTS data, which have to be dealt with separately. It contains 11 tables holding data and three lookup tables with information about statistical rectangle and areas (e.g. round fish area in the North Sea and sub-division in the Baltic Sea) and the relation between the different area types and statistical rectangles. The lookup tables for area and statistical rectangle can be used for both haul and area/stat.rec. related data. In addition there will be a number of other lookup tables such as ship, country and gear codes not shown in the figure.

The main table is the tblHaul that contains haul related mandatory data. The haul table is linked to the tables tblGear and tblAncillary. These tables contain optional data related to each haul and the tblHaul is therefore related to them with a one to one or zero relationship. To retrieve information on the different kinds of area tblHaul is combined with a lookup table for statistical rectangle, which is related to tlkpArea with a many-to-many relationship, as there might be more than one area type for each statistical rectangle and each statistical rectangle will be part of more than one area type (e.g. round fish area and the BTS otoliths sampling area covers the same statistical rectangle).

The cod index in the Baltic Sea (**and BTS + French/Spanish Surveys?**) is calculated by depth strata of 20 meters. These strata are also often used when extracting and analysing Baltic cod data in general. To meet these requests the table tblHaul is linked to a table with depth stratum for each haul in the Baltic Sea. As depth stratum is not needed for the IBTS data the relationship is a one to one or zero.

The table tblSpecies holds all the species related information and as there can be several or no species in a haul the tblHaul is related to tblSpecies with a one to many or zero relation. For each species there can be both ALK or length distribution information and the tblSpecies is therefore linked with a one to many or zero to the tables tblALK (containing ALK information) and tblLength (containing length distribution information calculated per hour). As it should be possible to deliver the raw length distribution instead of length distribution per hour tblLength is related to tblRawLength with a one to one or zero relationship

In some surveys it is accepted that ALK data are only provided by area (e.g. round fish area or statistical rectangle in North Sea IBTS) with no relation to the hauls at all. Also the data can be provided by either statistical rectangle or e.g. round fish area. This complicates the database structure slightly. Because ALK can be given by statistical rectangle or by area the data will be stored in two tables, one for statistical rectangle and one for area defined data. The two tables are then combined through the statistical rectangle (tlkpStatRec) and area (tlkpArea) tables.

In the BITS survey the fish weight in the ALK sub samples is reported. As this is only allowed in BITS and as ALK only can be submitted by haul in BITS there is a one to one or zero relationship between the table tblALK and tblMeanWeight.

This database design should make it possible to hold all data from these trawl surveys in the North-East Atlantic. However, data from the same area but from different surveys can/should not be aggregated when extracting the data and each survey therefore has to be identified by its name in the database. For this purpose a field called Survey is included in table tblHaul and in the area defined ALK tables.

3 DATAFLOW

Figure 2 shows the data flow for calculating number at age. First the age/length key (ALK) is aggregated by year, quarter and area and combined with number per length per haul and thereby calculating number at age per length per haul. From number at age and length per haul the indices and number at age and length per hour can be calculated. When calculating number at length per hour zero catch hauls are taken into account. From number at age and length per haul all other aggregations

levels can be extracted. The figure suggests that number at age and number at age and length per haul, statistical rectangle and area are the standard extractions from the database.

Number at age and length only needs to be calculated when the database is updated with new surveys or if data are resubmitted due to e.g. errors found after prior submission. This gives the possibility to combine the age data with the database as a new table (tblAge) linked to the species table (Figure 3) and thereby giving the user the possibility to combine the age data with all haul and species related information for haul-based extractions.

To make the extraction as flexible as possible it is suggested that date is included in the tables holding aggregated data. This means that data do not necessarily have to be extracted by quarter even if it is aggregated by area, as it is the case in both the IBTS and IBTS database at the moment. However, ICES data policy prescribe that public available data have to be aggregated per quarter.

In the flow diagram (figure 2) it is suggested that number at sex and maturity and number at sex is calculated as a multiplication of respectively the maturity/sex/age key and the sex/age key with number at age per area. In addition, it should be possible to extract maturity and sex at length and sex at length keys for further analysis. When calculated by haul the table can be combined with the species table in the database and thereby make it possible to extract other haul related data in combination with the maturity or sex data (Figure 3).

Hans Lassen and Lena Larsen tried to make maturity/sex at length keys per area to calculated number of mature at sex and length for Norway Pout. However, there are insufficient data to make the routine they used to a standard routine for the database. This might be different for other species, however, to implement this kind of routine in a database as a standard calculation for creating standard output one have to be sure that the routine will run and give reliable results for all standard species in all areas. Unfortunately, it doesn't seem that the data are sufficient.

4 OUTPUT AND INTERNET ACCESS

Figure 4 shows the suggested standard output tables from the database. All standard extractions will be accessible from the Internet. The figure shows that a password is needed to access haul related data while certain aggregated data are freely accessible. The figure also shows the future aspect of the database where it should be possible to merge all standard data with other databases in ICES such as the oceanographic data.

For the haul-based data it will be possible to select and include all or some of the haul related data such as gear coding and environment parameters. Figure 5 is an example of an Internet application where the users can click the parameters they want to include in a table with CPUE at length per haul. Depending on the time available for creating the web application it should also be possible to select the requested data on a map. As figure 4 shows, it will be possible to extract length, sex, maturity and age data on a haul basis.

The data will be held by date and for people without full access to the database, data will have to be aggregated over at least one month as described in the previous section. The aggregated tables will be predefined as the number of parameters are reduced when the data are aggregated.

When deciding exactly what and how output should be the output from IBTS and BITS databases should be taken into consideration. The standard tables from the BITS database can be found in Appendix I. Besides these tables the BITS database provides overview tables on number of hauls per year, quarter, ship and gear, number of hauls per year, quarter and haul duration, number of species per year, quarter and species recording code and number of otoliths per year, quarter and species. In addition, BITS provides maps from ArcView of distribution of relevant age classes by haul and number of cod by age group 1, 2 and 3+ per sub-division.

In IBTS a number of standard outputs can be extracted from the SIR database (Appendix I). However, they have very seldom been requested. The most often requested data from IBTS are the files called CHA and CHL (the former RIVO and HNR files) these are almost identical with the CHA and CHL tables in BITS (Appendix I). IBTS also provide maps with number of daytime hauls, total number of hauls and number at group 1, 2, 3+ for the standard species, which is published in the quarterly IBTS survey report.

The BTS have defined a number of outputs to be presented for plaice and sole for each survey BTS consisting of (ICES CM 2001/G:06 ref: ACFM,E):

- Abundance indices per year
- Mean length-at-age per year

For the surveys that are part of the database (e.g. the offshore beam trawl surveys) the following results should be presented:

- Map showing the area in ICES rectangles covered by each survey in the most recent year;
- Map showing for all surveys combined the number of hauls per ICES rectangle in the most recent year and over all years;
- Figure of all fish species caught per survey ranked according to their overall abundance. Results should include both the mean abundance per haul as well as the number of hauls in which the species occurred;
- Table with catches of the 26 most abundant fish species per management area, per year, for the last ten years;
- Maps showing the distribution of the 26 most abundant fish species. Shown are the catches per ICES rectangle in the most recent year and mean catch over all years;
- Figure of all the second reporting level (see chapter 5) benthic species caught per survey ranked according to their overall abundance. Results should include both the mean abundance per haul as well as the number of hauls in which the species occurred;
- Table with catches of all the first reporting level benthic species per management area, per year, for the last ten years;
- Maps showing the distribution of all the first reporting level benthic species. Shown are the catches per ICES rectangle in the most recent year and mean catch over all years.

To allow comparison between surveys and areas all catches should be expressed as numbers per eight meter beam trawl per hour trawling.

5 DATA EXCHANGE

The exchange formats defined in the BITS and IBTS manuals are specified as fixed width text format and data output from the IBTS and BITS database have been either as fixed width text or Excel files (BITS only). The DATRAS database should accept and provide files in these formats. However, there has in the later years been an increased focus on the XML format and to ensure the database for the future and increase the flexibility of data format and the database XML should be considered as an additional possibility for data exchange.

To change the exchange format into a XML format would only be reasonable if the checking of the data also will be done within XML (XML-schemas/DTD). However, as Neil Holdsworth describes in the report Data Check Program this technology is not fully developed yet and data checking in XML (DTD/schemas) are not working in such a way that it can be used in a checking program.

XML files with data output could very easily be created from the database. It is expected that different laboratories might request this format, as XML becomes the main data exchange format in the Microsoft environment in the future.

6 CHECKING PROGRAM

A part of the DATRAS project is to create a checking program, which should be able to handle data for all the surveys in the database. At the moment two checking programs exist for trawl data in ICES; one for the IBTS and one for the BITS data. The BITS checking program is almost a copy of the IBTS checking program.

The checking programs are written in COBOL and every time it has to be updated one has to modify the code, update it, compile the code, upload the new version to the Internet for the user to download it and run it in DOS. A major problem with these programs is therefore that different versions of the checking program are used at the different institutes and that updating the program with e.g. a new species code takes a lot of effort as one has to go into the code and find where the update should be done.

To overcome these kind of problems ICES would like to make an Internet based checking program for the future with a database behind. To update simple codes ICES would only have to enter the table holding information on e.g. species and make the update there and the update will automatically be reflected at the front end.

The checking program will be made in collaboration with the Environment Department in ICES to create one standard for future checking program in ICES. The report Data Check Program describes different possibilities of how such a program can be developed and gives a recommendation on how ICES would prefer to do it.

Appendix III shows the checks performed in the current BITS checking program, which are almost identical with those done by the IBTS checking program. The only differences are in ranges, values and areas. The future checking program will include the same checks but beside these the DATRAS Steering group may come up with new requests. E.g. the IBTS meeting 2001 suggested that checks on rare species should be included/extended.

7 CALCULATING CPUE AT AGE AND LENGTH

A detailed description on calculations in the BITS and IBTS database can be found in I V and V respectively. A detailed description on how and what to calculate for the newly available data will have to be written by the steering group members in the nearest future.

Two different approaches have been used for calculating the CPUE at age and length in the IBTS and BITS database respectively. The following section describes the difference in the two approaches.

The approaches differ in how the age/length key (ALK) are generated. The direct observation is a distribution of fish ages for given length class and both procedures start by summing the ALKs over the categories agreed (depth/area/time). However, the sampling procedure does not ensure that there is an age distribution for each length class. These so called holes (length classes where CPUE > 0 but without an age composition) should then be filled in and it is these fill-in procedures that differ. In BITS the solution has been to weight ALK with mean CPUE per length and smooth the weighted data with lowess smoother over 5 point.

In IBTS it is first decided if the age/length is sufficient or to 'thin'. If not, keys will be borrowed from the neighbour areas. Even when the key is sufficient (with or without substitutions keys) the key might have 'holes'. These 'holes' are filled in with values from neighbouring length classes or if data are missing below a lower or above an upper limit, data are either allocated to zero/one year group or the age+ group.

The BITS approach is working very well on the Baltic cod, for which it was developed and tested. However, when using it on Norway Pout the procedure failed as each age group of Norway Pout only covers a few length classes. A loless smoother over 5 length classes does therefore not reflect the rapid change in age with length correctly. Thus, the BITS solution cannot be used for all species in the DATRAS database.

The IBTS procedure of ‘filling in the holes’ has been used for almost two decades now and seems rather robust. However, this procedure is partly manual, as borrowing ALK from a neighbouring area is not formalised. Also filling in holes will not reflect growth correctly something that becomes critical if there are larger holes in the ALK.

Defining a common strategy for use in all cases requires a research project on its own something that is not stipulated within the DATRAS project. The approach to such a project should be through a population model including growth, length dependent mortality and recruitment.

7.1 ALK AGGREGATION AREAS

ALK is aggregated over larger areas. In IBTS North Sea the area used is the roundfish area; in IBTS western - and southern shelf area are XXXX. BTS uses the Otoliths areas in the North Sea and XXX in the other management areas. In BITS the aggregation areas are the sub-divisions.

We need a complete list/overview of ALK areas.....

8 INDICES

The three different surveys groups (BITS, IBTS and BTS) provide indices for different species for different assessment working groups. Table 1 gives an overview of the different species, where they are caught and in which working groups the indices are used.

Table 1: Overview of the species, for which the surveys provide indices.

Survey	Area	Species	WG
BTS	North Sea, Celtic Sea and Irish Sea	plaice	northern shelf WG (VIIa), southern shelf (VIIe, VIIe, VIIg,f), and North Sea and Skagerrak WG
	North Sea, Celtic Sea and Irish Sea	sole	northern shelf WG (VIIa), southern shelf (VIIe, VIIe, VIIg,f), and North Sea and Skagerrak WG
IBTS Q1	North Sea and Skagerrak	cod	North Sea and Skagerrak WG
	North Sea and Skagerrak	whiting	North Sea and Skagerrak WG (Usually but not in 2001)
	North Sea and Skagerrak	saithe	
	North Sea and Skagerrak	haddock	North Sea and Skagerrak WG
	North Sea and Skagerrak	norway pout	North Sea and Skagerrak WG
	North Sea and Skagerrak	herring	Herring Assessment south of 56 WG
	North Sea and Skagerrak	sprat	Herring Assessment south of 56 WG
	North Sea and Skagerrak	mackerel	Mackerel and Horse Mackerel Assessment WG (Made available–not used but deemed important for distribution etc)
IBTS Q3	North Sea and Skagerrak	Cod	North Sea and Skagerrak WG (Scottish & English separately)
	North Sea and Skagerrak	Whiting	North Sea and Skagerrak WG (Usually but not in 2001)
	North Sea and Skagerrak	Haddock	North Sea and Skagerrak WG (English data only)
	North Sea and Skagerrak	Norway pout	North Sea and Skagerrak WG (Scottish & English separately)
IBTS Q1	VIa	Cod	Northern Shelf WG
	VIa	Haddock	Northern Shelf WG
	VIa	Whiting	Northern Shelf WG
	VIa	Monkfish	Northern Shelf WG (Exploratory data analysis)
	VIa	Mackerel	Mackerel and Horse Mackerel Assessment WG (Made available–not used but deemed important for distribution)

IBTS Q4	Vla	Mackerel	etc) Mackerel and Horse Mackerel Assessment WG (Made available–not used but deemed important for distribution etc)
BITS	Baltic Sea	cod	Baltic WG
EVHOE	VIIghj	Cod	Southern Shelf WG
Q4	VIIghj	Whiting	Southern Shelf WG
	VIIghj, VIIab	Hake	Southern Shelf WG
	VIIghj, VIIab	Megrim	Southern Shelf WG
	VIIghj, VIIab	Lophius spp	Southern Shelf WG
	VIIghj, VIIab	Mackerel	Mackerel and Horse Mackerel Assessment WG (Made available–not used but deemed important for distribution etc)

8.1 BEAM TRAWL SURVEY

The BTS survey consists of many surveys that is not totally integrated and in different areas. The target species are flatfish and indices are calculated for plaice and sole. As the surveys are not completely integrated the indices are calculated separately for each survey/area.

In Division VII, area stratified age–length keys are used to derive mean numbers at age by rectangle within the survey area. The index of abundance is derived from the unweighted arithmetic mean across all rectangles (ICES CM 2001/G:06). In the North Sea, Netherlands BTS, the indices are calculated as above except there is used a reduced section of the full survey area known as the “Index Area”.

8.2 BITS INDICES

In the BITS database the cod indices are calculated for the two areas sub-division 22+24 and 25+26+28. First the mean age at length per depth stratum and sub-division (where CPUE at length per haul have been weighted by fishing power for each research vessel) are calculated and weighted with m² of the stratum. From these means mean at age per sub-division is calculated and then the mean per index area. A more detailed description of how the Baltic cod indices are calculated can be found in Appendix IV and the fishing power used to weight CPUE with in table 3.

Table 3: Fishing power BITS
Calculated by the Baltic Fisheries Assessment Working Group
(ICES CM 1998/ACFM:16)

Ship	Gear	Species	Fish power
ARG	FOT	Cod	1
ARG	GOV	Cod	0.948443
ATL	HAK	Cod	0.885257
ATLD	HAK	Cod	0.885257
BAL	P20	Cod	0.323406
BPE	LBT	Cod	0.419338
BPE	LPT	Cod	0.419338
CLV	LBT	Cod	0.419338
DAN2	GRA	Cod	0.537068
DAN2	GRA060	Cod	0.537068
DAN2	GRA110	Cod	0.537068
GDY	P20	Cod	0.323406
MON	DT	Cod	0.419338
MONL	DT	Cod	0.419338

SOL	CHP	Cod	0.553931
SOL	H20	Cod	0.553931
ZBA	LBT	Cod	0.419338
ZBA	LPT	Cod	0.419338

8.3 IBTS INDICES

In IBTS North Sea indices are calculated by species dependent index area for eight species (herring, sprat, mackerel, cod, haddock whiting, norway pout and saithe). The indices are calculated as mean at age per statistical rectangle and then as a mean of the statistical rectangles over the index area. Some statistical rectangles are reduced in size due to land or very shallow water. For herring and saithe the mean CPUE at age are weighted with the percent covered with water for these statistical rectangles. Whatever mean CPUE also should be weighted for the other standard species will be analysed during the winter 2002 and decided on the next IBTS meeting.

The index areas are shown in various working group reports from the IBTS group. However, the IBTS working group intends to establish new indices areas for cod (eastern channel; Div. VIId), North Sea (sub-area IV and Skagerrak), whiting (eastern Channel and North Sea) and haddock and norway pout (North Sea; sub-area IV and Skagerrak/Kattegat; Div. IIIA).

IBTS indices for western and southern divisions... This information has to be provided by the steering group.

9 SPECIES CODES

In the IBTS and BITS exchange formats and in the BITS database the 10 digits NODC codes are used as species codes while in the IBTS database Rubin codes are used as codes. Neither of the two systems are updated any longer, however, the National Oceanographic Data Center has come up with a new code system. In the last update of the old NODC codes (version 8) they wrote:

Up to now the NODC Taxonomic Code was based on a system of code numbers that reflected taxonomic relationships. Over the years the problems and shortcomings created by this approach have become more apparent. At the same time, relational database management systems have been developed that enable efficient searches and retrievals to be conducted without the need for taxonomic information to be built into the codes. Therefore, Version 8.0 of the NODC Taxonomic Code is a transition product that includes both the original types of codes as well as codes for a new system that will overcome past problems and provide greater capabilities.

And later:

This version of the NODC Taxonomic Code contains these "intelligent" code numbers, but also introduces non-intelligent Taxonomic Serial Numbers (TSNs) that will be used in future editions of the NODC Taxonomic Code. Therefore, this version of the NODC Taxonomic Code is the last version of the NODC Taxonomic Code that contains new "intelligent" code numbers. The NODC will preserve and publish the old "intelligent" NODC code numbers for the foreseeable future. The contents of this CD-ROM show the final state of the old code numbers however, and they will not be deleted or modified, nor will new ones be assigned.

As new taxa are added to the file they will be assigned Taxonomic Serial Numbers. The TSN remains with the taxon to which it was assigned, even though the taxon itself may go in or out of synonymy or be entirely reclassified. This system allows the inclusion of multiple levels of taxonomy and greatly improves the overall management of the NODC Taxonomic Code.

The question is whether ICES should adapt a new code system or keeps the old NODC code system in the exchange format. To change the system could cause errors in the rarer species and it will need an

agreement from all the involved institutes, as they have to make changes to their systems. On the other hand in the BITS data there have already been problems with missing codes and BITS have had to invent their own codes.

In the database any kind of system could be implemented, even the Latin names. However, the system selected doesn't have to be the same as the one the user sees when searching and extracting data. Ideally the user can search on both Latin and English names and e.g. NODC codes. This is just a question of having a key table, which translate from one name to another. The system selected in the database does therefore not have to be the same as the one used for searching.

10 CONCLUSION

This document shows that a number of things must be sorted out and documented before programming can start. The following list is a summary of issues encountered when writing this report:

- Create manuals for all data, which will be included in the DATRAS database, as in IBTS and BITS
- Update list of data checks per survey
- Compile a paper on how to calculate abundance indices which the database should be able to provide (as in Appendix IV for BITS)
- Agree on what to do with sex, maturity and weight
- Agree on standard output from the database
- Agree on species codes in database and in exchange format
- Agree on search facilities: should it be possible to search on e.g. Latin and English names and would it be acceptable to extract aggregated data over one month time period?
- List of areas for the different surveys
- Define index area for IBTS in previous time and present
- Should it in the future only be possible to deliver ALK per haul?
- Sub-samples delivery for length data (HL-records)
- Other new variables in the database

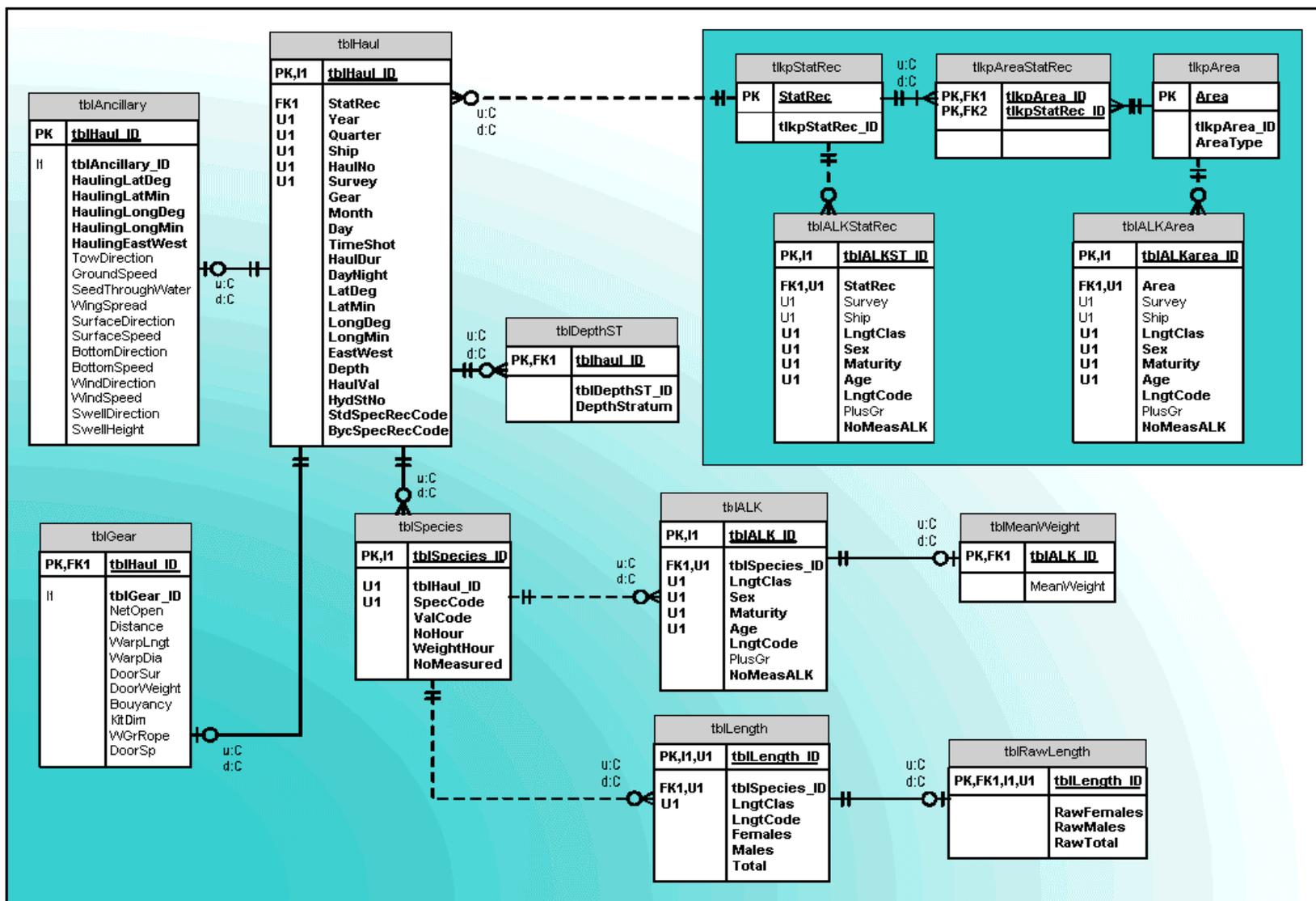


Figure 1. DATRAS database structure.

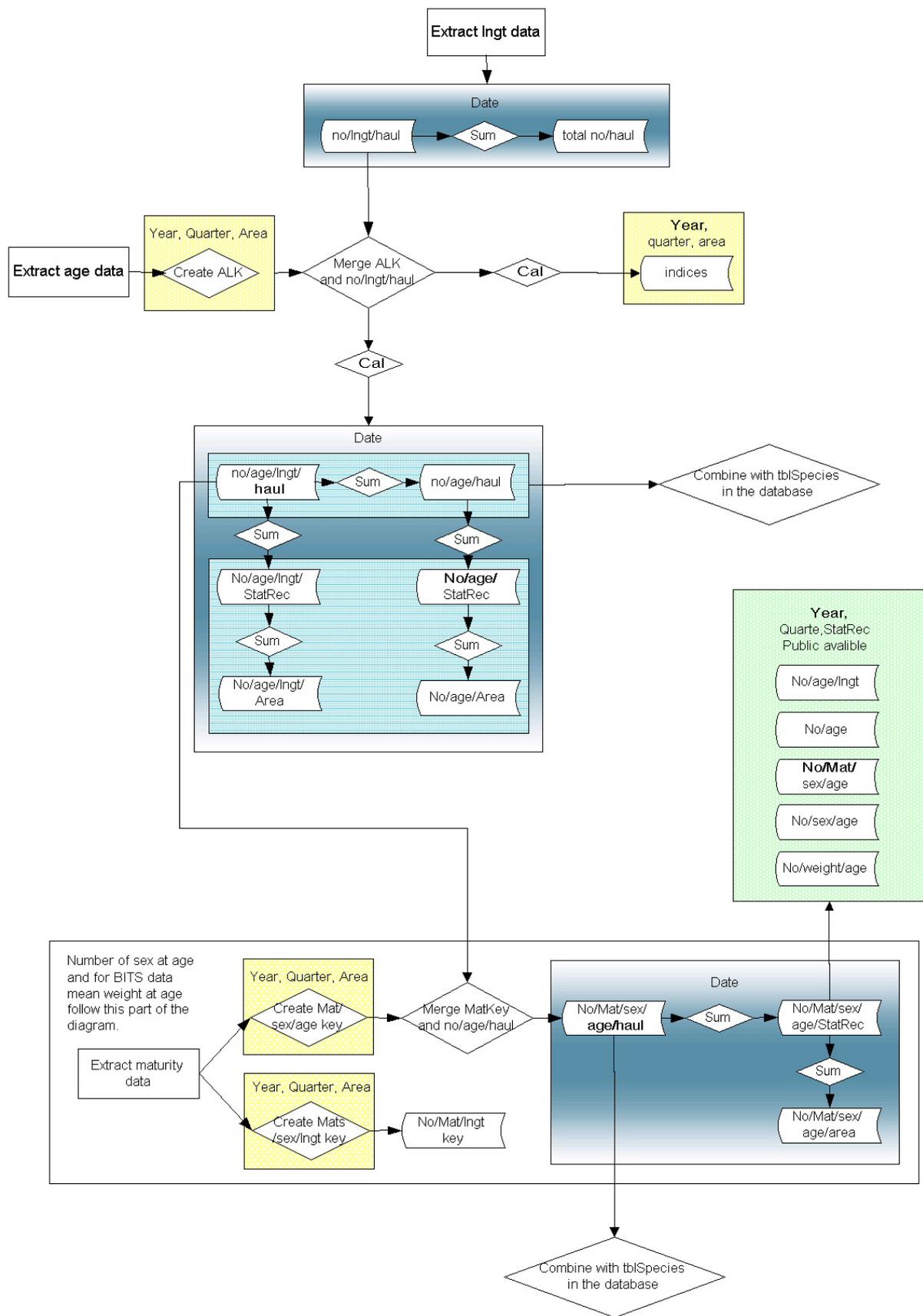


Figure 2. Dataflow in the DATRAS database

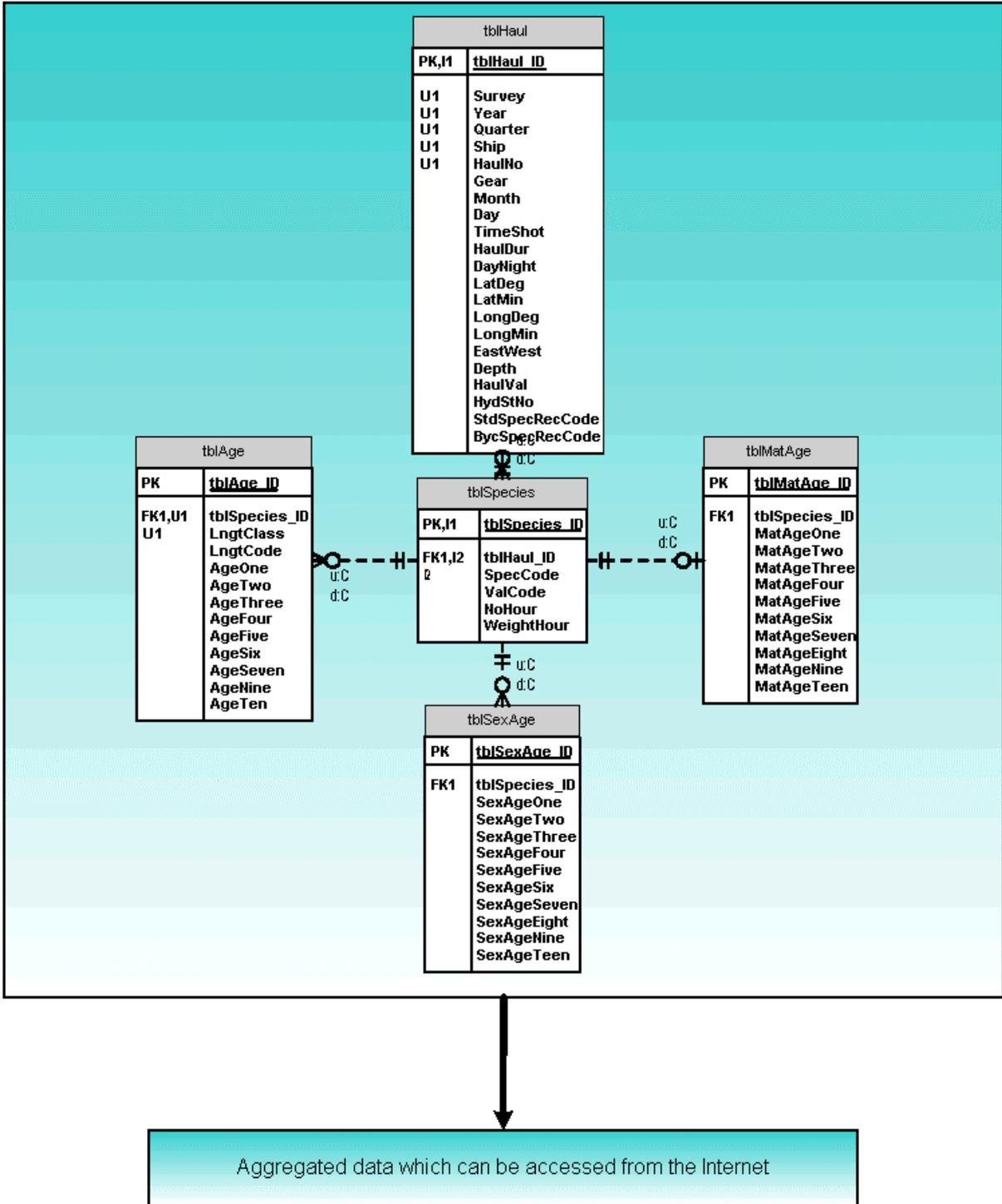


Figure 3. Suggested database structure for calculated age and maturity.

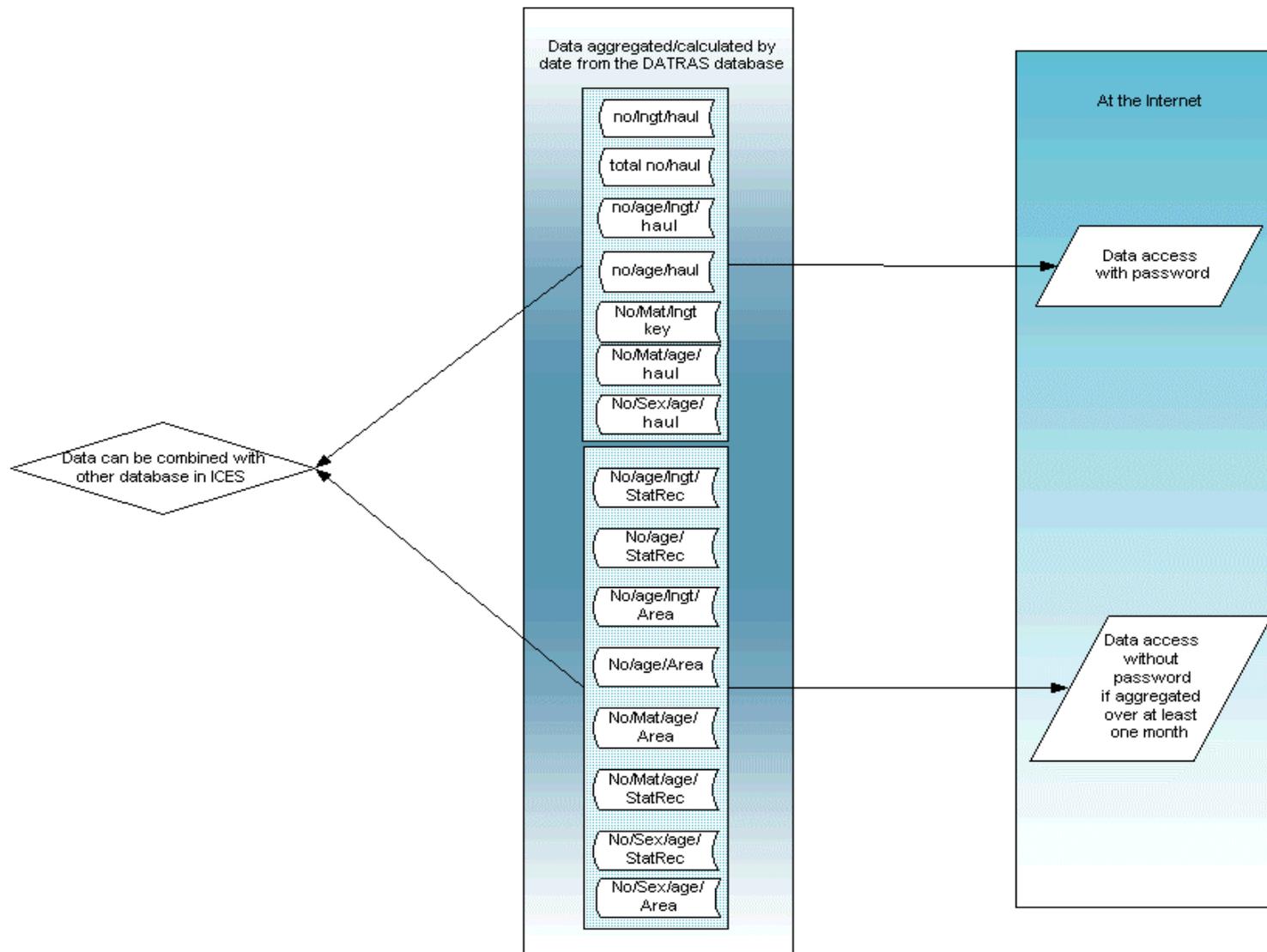


Figure 4. Standard output and connection to the Internet.

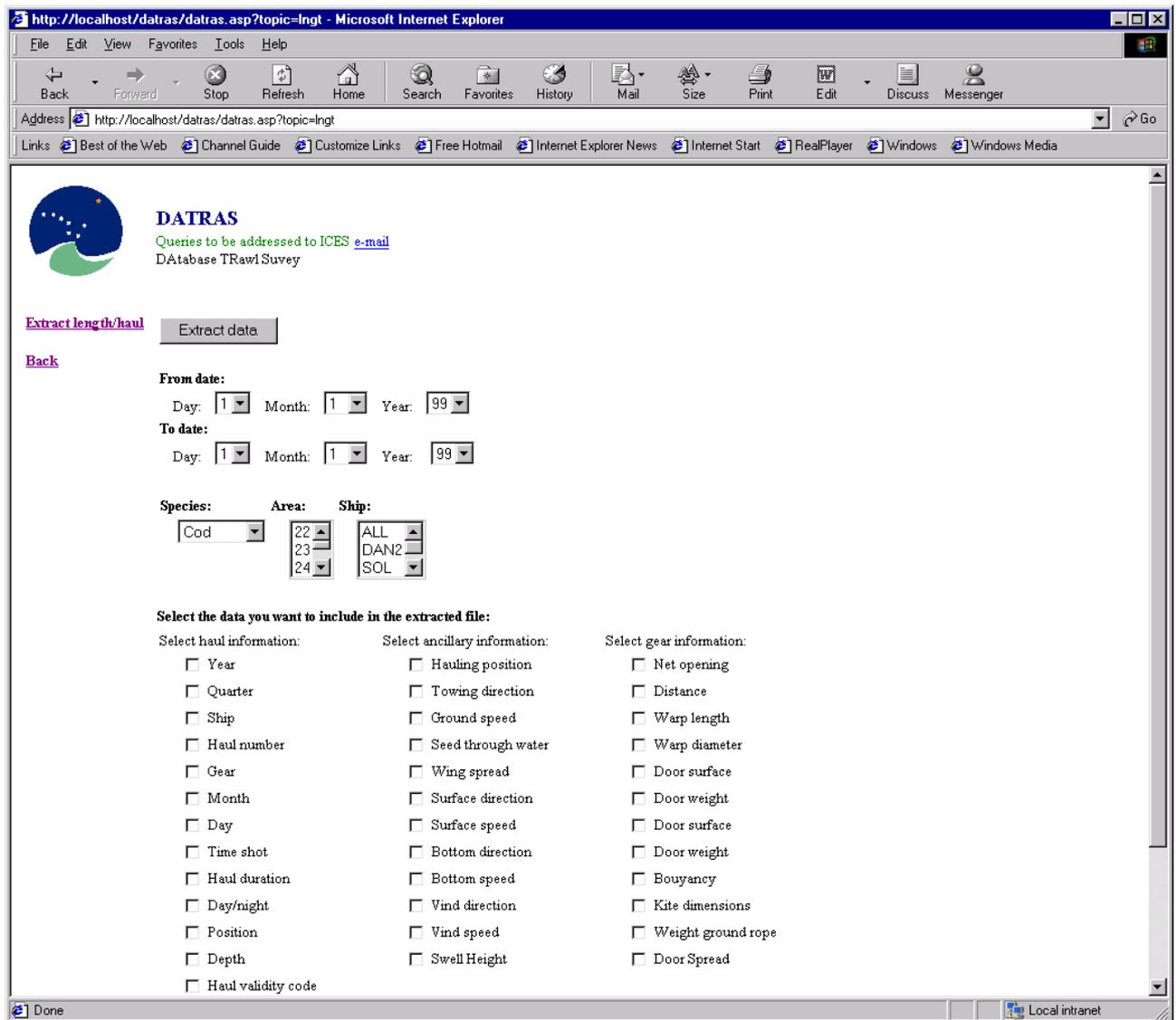


Figure 5. The first draft on how data extraction for the haul related data could be done.

APPENDIX I: STANDARD TABLES IN THE BITS AND IBTS DATABASE

BITS: All standard tables generated in the database can be exported as fix width text files.

ALK_ST_Lngt	ALK_ST_Sum	ALK_SD_Lngt	ALK_SD_Total	MaturityOgive	MeanWeight	MeanCPUE_ST_Lngt	MeanCPUE_ST_Total	MeanCPUE_SD_Lngt	MeanCPUE_SD_Total
Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Sub-division	Sub-division	Sub-division	Sub-division	Sub-division	Sub-division	Sub-division	Sub-division	Sub-division	Sub-division
Depth stratum		Depth stratum	Depth stratum	Depth stratum	Depth stratum				
Species code	SpecCode	Species code	Species code	Species code	Species code				
Length class		Length class				Length class		Length class	
Plus group	Plus group	Plus group	Plus group	Age	Age				
Mean CPUE		Mean CPUE	Mean CPUE	Mean CPUE	No. obs.	Mean CPUE	Mean CPUE	Mean CPUE	Mean CPUE
Age null	Age null	Age null	Age null	No immature	Mean weight				
Age one	Age one	Age one	Age one	No mature					
Age two	Age two	Age two	Age two	Fraction mature					
Age three	Age three	Age three	Age three						
Age four	Age four	Age four	Age four						
Age five	Age five	Age five	Age five						
Age six	Age six	Age six	Age six						
Age seven	Age seven	Age seven	Age seven						
Age eight	Age eight	Age eight	Age eight						
Age nine	Age nine	Age nine	Age nine						
Age ten	Age ten	Age ten	Age ten						

BITS con.

CHA/CHL

Year
Quarter
Ship
St. No
Haul No.
Gear
Months
Day
Time shot
Haul duration
Day/night
Sub-div
Depth stratum
Depth
Lat. deg.
Lat. min.
Lon. deg.
Lon. min.
SpecCode

CHA		CHL
Plus gr.	76-78	Length class
Age null	79-84	No. obs.
Age one		
Age two		
Age three		
Age four		
Age five		
Age six		
Age seven		
Age eight		
Age nine		
Age ten		

IBTS: Standard output from the SIR database

Standard tables	Data
Number of otoliths sampled per area	Length class age group number of otoliths
Age/length keys per area	Length class age group number of otoliths
Substitution age/length keys	Length class age group number of otoliths
Age/length keys, incl. substitution keys	Length class age group number of otoliths
Standard and by-catch species overview	No. hauls no. rect total no. per hour per haul mean no. per hour per haul st.dev. no. per hour per haul min no. per hour per haul max no. per hour per haul
Summary information per std. species	Sampling area number of rect. number of valid hauls per sampling area total no. per hour per haul per sampling area mean no. per hour per haul per sampling area number of valid hauls per standard area total no. per hour per haul per standard area mean no. per hour per haul per standard area
Length distribution by area	Length class No. per length per standard area No. per length per sampling area Mean length per standard area Mean length per sampling area St. dev. per standard area St.dev. per sampling area Min length class per standard area Min length class per sampling area Max length class per standard area Max length class per sampling area
Relative length dist. (%) by sampling area	Length class Relative total catch per length Relative catch per length per sampling area
Age/length dist. by sampling area	Length class Age group No. per length and age Mean per length and age St. dev. per length and age Min per length and age Max per length and age
Age dist. by stat. rec. number per hour per haul	No. valid hauls per stat. rec. No. at age per stat. rec.
Relative age dist. (%) by stat. rec.	No. valid hauls per stat. rec. No. at age per stat. rec.
Mean length and st. dev. by age group and stat. rec.	Mean length per age and stat. rec. St. dev. per age and stat. rec.
Min/max length by age group and stat. rec.	Min length per age and stat. rec. Max length per age and stat. rec.

Age distribution by area, sum of all rect.	Total no. at age per standard area
	Total no. at age per sampling area
Mean number per hour per haul per stat. rec. by age group and area	Mean CPUE per standard area
	Mean CPUE per sampling area
Relative age dist (%) by area	Relative no. at age per standard area
	Relative no. at age per sampling area
Mean length and st. dev. by age group and area	Mean no at age per standard area
	St. dev. per standard area
	Mean no at age per sampling area
	St. dev. per sampling area

APPENDIX II: XML VALIDATION

by Neil Holdsworth

Validation of XML Files

In the realm of XML there are two commonly repeated phrases that are distinctly separate; Well-Formedness and Validation.

Well-Formedness

Means to check an XML file to determine whether every start tag element has a matching end tag, that no illegal characters have been used to identify tag elements etc.

Validation

Takes an XML file and checks the content of it according to a set of rules provided by an XML Schema or DTD.

When an XML file is parsed (to a browser for instance) the file will always be checked for well-formedness but it may not necessarily be validated.

Therefore, we would have to use a data schema to validate our XML data files.

Assuming that I want my XML to be validated by a schema I have created how can I report the errors in my XML file?

There are essentially two ways to report the errors from an XML document; The Document Object Model (DOM) and the Simple API for XML (SAX).

DOM and SAX are both interface objects that hold whatever XML file we request into it and allow us to manipulate the object with a set of predefined functions and events. The fundamental difference is how these two methods load an XML file.

Document Object Model (DOM)

The DOM object takes the whole XML file and loads it in and then makes the Object available to the user/program for manipulation.

Simple API for XML

The SAX object reports events back to the user/program as it is loading and is not constrained by waiting for the whole file to load.

Thus, if we were looking for a particular value in an XML file with 1000 elements(records) and that value was in the 3rd record of the file, SAX would return the value much faster than DOM.

Errors in DOM

Because of the way DOM loads an XML file the DOM object only gives us the possibility of reporting on the last error it encountered while parsing the XML file. This means I would have to correct the reported error in the XML file and then re-parse the file to see if any other errors occurred and so on until all the errors were

found. In a normal situation where we would be dealing with a large XML data file and more than 2-3 errors this method is completely unworkable.

Errors in SAX

Because SAX is event driven it should be possible to catch each error as it occurs in the load and then continue parsing the document from where the error occurred. Unfortunately the maxim “the gap between theory and practice is not as wide in theory as it is in practice.”

SAX is primarily a set of Java Classes, but because my knowledge of Java is limited I have opted to look at the Microsoft implementation of SAX (XML Parser 4.0 version) which is compatible with VB and VBScript.

There are two areas of interest in the MS SAX interface, the ErrorHandler and the ReaderControl.

ErrorHandler

This interface is where we would handle an error from the XML file been parsed. Despite there been 3 methods associated with the ErrorHandler (Error, Warning and Fatal Error) only FatalError is implemented ie. Usable. Unfortunately, fatal errors are non-recoverable, they stop the parsing of the XML file and thus the first error we encounter will stop the parsing of the file and although we can do many things at this point we can't continue parsing from that error point.

ReaderControl

This interface works in conjunction with the SAX reader interface and allows the parse operation to be Aborted, Suspended or resumed upon a conditional event. Once more, the event that we would like to trigger this interface is the Error from the errorhandler but the Fatalerror condition aborts the parse before we can get our hands on it. Furthermore, this interface is a non-standard add on which means it shouldn't be relied upon.

At this point, it is clear that the Microsoft implementation is a partial implementation of error handling. To further complicate matters, SAX is not a standard in the way XML 1.0 is but is just a collection of resources put together by a group of Java developers. Which makes it very hard to determine what SAX should be capable of (there is no Documentation) and whether Microsoft will implement it. Abandoning the Microsoft parser would entail implementing Java. From what I have seen so far it would seem the Java classes allow the reporting of errors and warning as well as fatal errors, however I'm not prepared to take this issue any further for now.

A few more words about SAX can be found at <http://sax.sourceforge.net/>

The MS implementation of SAX can be found at <http://msdn.microsoft.com/library/>

APPENDIX III: CHECKS PERFORMED BY THE PRESENT CHECKING PROGRAM IN THE BITS AND IBTS DATA

IBTS Fishery trawl data

	<i>start</i>	<i>fieldcode description</i>	<i>error</i>
HH	Haul Information		
	1	RecordType Rec. type must be HH	error
	3	Quarter Quarter = 1, 2, 3 or 4 Quarter depend on month	error
	4	Country Check valid list code	error
	7	Ship Check valid list code Ship depend on country	error
	11	Gear Check valid list code Gear depend on ship	error
	27	HaulNo A haul may only appear once in a set of HH-records =>1 and <= 999	error
	30	Year =<65 and <=99 or =>00 and <=40	error
	32	Month =>1 and <=12	error
	34	Day =>1 and (<=28 or <= 29 or <=30 or <=31) - dep. on month and year	error
	36	TimeShot =>1 and <=2400	error
	40	HaulDuration =>5 and <=150	error
	43	Day/night D, N, Space Day/night is coded according to the description of day/night in the BITS manual (depend on time and position)	error
	44	LatDeg =>53 and <=66	error
	46	LatMin =>0 and <=59	error
	48	LonDeg =>11 and <=31	error
	50	LonMin =>0 and <=59	error
	52	East/West E	error
	53	Depth =>10 and <=150, spaces, =>5 and <=150 in sub div 22 + 24	error
	56	HaulValidity I, V, N	error
	65	SpeciesRecordingCode Check valid list code	error
	67	Netopening =>15 and <=100, spaces	error
	70	Distance =>1850 and <=9999, spaces	error
	74	Warplength =>100 and <=999, spaces	error
	77	Warpdiameter =>10 and <=60, spaces	error
	79	DoorSurface =>10 and <=100, spaces	error
	82	DoorWeight =>50 and <=2000, spaces	error
	86	Buoyancy =>50 and <=200, spaces	error
	90	KiteDimensions =>5 and <=20, spaces	error
	92	WeightGroundRope =>0 and <=800, spaces	error
	96	DoorSpread =>25 and <=200, spaces	error
HE	Haul Information Ocean		
	3	Quarter Records referring to the same haul must contain identical values for Quarter	error
	4	Country Records referring to the same haul must contain identical values for Country	error
	11	Gear Records referring to the same haul must contain identical values for Gear	error
	21	StationNumber Records referring to the same haul must contain identical values for StatNo	error
	41	Towing Direction =>0 and <=360, spaces	error
	44	GroundSpeed =>20 and <=60, spaces	error
	46	SeedThroughWater =>10 and <=99, spaces	error
	48	WingSpread =>12 and <=30, spaces	error
	50	SurfaceCurrentDirection =>0 and <=360, spaces	error
	53	SurfaceCurrentSpeed =>0 and <=100, spaces	error
	56	BottomCurrentDirection =>0 and <=360, spaces	error
	59	ButtomCurrentSpeed =>0 and <=100, spaces	error
	62	WindDirection =>0 and <=360, spaces	error
	65	WindSpeed =>0 and <=100, spaces	error
	68	SweelDirection =>0 and <=360, spaces	error
	71	SweelHeight =>0 and <=999, spaces	error
HL	Length frequency distribution		
	1	RecordType Rec. type must be HL If HH has a HL; If year, ship, haul no and species code are the same as previous error recorded then quarter, country, gear, rig, doortype and st. no must also be the same There always have to be a HH record for each HL record Not more than one record per species, length class and sex are allowed	error
	3	Quarter Records referring to the same haul must contain identical values for Quarter	error
	4	Country Records referring to the same haul must contain identical values for Country	error
	11	Gear Records referring to the same haul must contain identical values for Gear	error

21	StationNumber	Records referring to the same haul must contain identical values for StatNo	error
32	SpeciesCode	Check valid list code	error
42	ValidityCode	If SpeciesCode = herring or sprat then LengthClassCode = ., 9 or 0 and if validity code is 09 then NoHour, CatchWeight, NoMeasured, LengthClassCode, MinLengthClass and HLNoAtLength = space	error
error		if validity code is 04 then NoHour > 0	error
		if validity code is 04 then LengthClassCode, MinLengthClass and HLNoAtLength = space	error
		if validity code is 00 then NoHour, CatchWeight, NoMeasured, LengthClassCode, warning	error
		MinLengthClass and HLNoAtLength = space	error
		Validity code must be compatible to haul validity code	error
		Records referring to the same haul must contain identical values for ValidityCode	error
		Only one record can be given per haul and species with the ValidityCode 4 or 9	error
44	NoHour	Check valid list code	error
		Records referring to the same haul must contain identical values for NoHour	error
		If NoHour = 0 then CatchWeight = 0 and visa versa, unless catch weight < 100g =>0 and <=9999999	error
51	CatchWeightHour	=>0 and <=99999, spaces	error
		Records referring to the same haul must contain identical values for CatchWeight	error
56	NoMeasured	=>0 and <=999, spaces	error
		Records referring to the same haul must contain identical values for NoMeasured	error
59	LengthClassCode	., 0, 1, 2, 5 or 9	error
		If LengthClassCode = 0 then MinLengthClass must be half cm units	error
		If LengthClassCode = 2 then MinLengthClass must be 2 cm units	error
60	MinLengthClass	If LengthClassCode = 5 then MinLengthClass must be 5 cm units and > 60	error
		If species, LengthClassCode and MinLengthCode are OK then max length is checked	error
		If LengthClassCode = 0 then MinLengthClass must be half cm units	error
		Species referring to the same haul must have the same LengthClassCode =>1 and <=999, spaces	error
63	HLNoAtLengthHour	=>1 and <=9999999, spaces	error
		For each species; NoHour = sum of HLNoAtLengthHour	error
69	Sex	M or F, spaces	error

CA SMALK

1	RecordType	If HH has a CA; If year, ship, haul no and species code are the same as previous recorded then quarter, country, gear, rig, doortype and st. no must also be the same	error
		Only on record can be given per haul, species, length class, sex, maturity and age group	error
3	Quarter	Records referring to the same haul must contain identical values for Quarter	error
4	Country	Records referring to the same haul must contain identical values for Country	error
11	Gear	Records referring to the same haul must contain identical values for Gear	error
21	StationNumber	Records referring to the same haul must contain identical values for StatNo	error
42	Area(SubDivRFish)	Check valid list code	error
		SubDiv and StatRec must be consistent and legal	error
44	StatRec	Check valid list code	error
52	LengthClassCode	If SpeciesCode = herring or sprat then LengthClassCode = ., 9 or 0 and Species referring to the same haul must have the same LengthClassCode	error
		If LengthClassCode = 5 then MinLengthClass must be 5 cm units and > 60	error
		If LengthClassCode = 2 then MinLengthClass must be 2 cm units	error
		If LengthClassCode = 0 then MinLengthClass must be half cm units	error
53	MinLengthClass	., 0, 1, 2, 5 or 9	error
		If species, LengthClassCode and MinLengthCode are OK then max length is checked	error
56	Sex	M or F, spaces	error
57	Maturity	=>1 and <=5, spaces	error
58	PlusGP	+, spaces	error
59	Age	=>0 and <=99, spaces	error
		If age = space then PlusGr = space	error
		Per haul, species and length class only the oldest age group can be PlusGR	error
61	CANumberAtLengthHour	=>1 and <=999, spaces	error
64	MeanWeight	=>0 and <=99999, spaces	error

IBTS Fishery trawl data

<i>start</i>	<i>fieldcode</i>	<i>description</i>	<i>error</i>
HH Haul Information			
1	RecordType	Rec. type must be HH	error
3	Quarter	Quarter = 1, 2, 3 or 4	error
4	Country	Quarter depend on month Check valid list code	error
7	Ship	Check valid list code Ship depend on country	error
11	Gear	Check valid list code Gear depend on ship	error
27	HaulNo	A haul may only appear once in a set of HH-records =>1 and <= 999	error
30	Year	=<65 and <=99 or =>00 and <=40	error
32	Month	=>1 and <=12	error
34	Day	=>1 and (<=28 or <= 29 or <=30 or <=31) - dep. on month and year	error
36	TimeShot	=>1 and <=2400	error
40	HaulDuration	=>5 and <=90	error
43	Day/night	D, N, Space Day/night is coded according to the description of day/night in the IBTS manual (depend on time and position)	error
44	LatDeg	=>50 and <=64	error
46	LatMin	=>0 and <=59	error
48	LonDeg	=>0 and <=15	error
50	LonMin	=>0 and <=59	error
52	East/West	E,W	error
53	Depth	=>10 and <=300, spaces	error
56	HaulValidity	I, V, P	error
65	SpeciesRecordingCode	Check valid list code	error
67	Netopening	=>25 and <=100, spaces	error
70	Distance	=>1850 and <=9999, spaces	error
74	Warplength	=>100 and <=999, spaces	error
77	Warpdiameter	=>10 and <=60, spaces	error
79	DoorSurface	=>10 and <=100, spaces	error
82	DoorWeight	=>50 and <=2000, spaces	error
86	Buoyancy	=>50 and <=200, spaces	error
90	KiteDimensions	=>5 and <=20, spaces	error
92	WeightGroundRope	=>0 and <=300, spaces	error
96	DoorSpread	=>50 and <=180, spaces	error
HL Length frequency distribution			
1	RecordType	Rec. type must be HL If HH has a HL; If year, ship, haul no and species code are the same as previous recorded then quarter, country, gear, rig, doortype and st. no must also be the same There always have to be a HH record for each HL record Not more than one record per species, length class and sex are allowed	error
3	Quarter	Records referring to the same haul must contain identical values for Quarter	error
4	Country	Records referring to the same haul must contain identical values for Country	error
11	Gear	Records referring to the same haul must contain identical values for Gear	error
21	StationNumber	Records referring to the same haul must contain identical values for StatNo	error
32	SpeciesCode	Check valid list code If SpeciesCode = herring or sprat then LengthClassCode = ., 9 or 0	error
42	ValidityCode	if validity code is 09 then NoHour, CatchWeight, NoMeasured, LengthClassCode, MinLengthClass and HLNoAtLength = space if validity code is 04 then NoHour > 0 if validity code is 04 then LengthClassCode, MinLengthClass and HLNoAtLength = space if validity code is 00 then NoHour, CatchWeight, NoMeasured, LengthClassCode, MinLengthClass and HLNoAtLength = space Validity code must be compatible to haul validity code	error error error warning
		Records referring to the same haul must contain identical values for ValidityCode	error
		Only one record can be given per haul and species with the ValidityCode 4 or 9	error
44	NoHour	Check valid list code Records referring to the same haul must contain identical values for NoHour If NoHour = 0 then CatchWeight = 0 and visa versa, unless catch weight < 100g	error
51	CatchWeightHour	=>0 and <=9999999 =>0 and <=99999, spaces Records referring to the same haul must contain identical values for CatchWeight	error
56	NoMeasured	=>0 and <=999, spaces	error

		Records referring to the same haul must contain identical values for NoMeasured	error
59	LengthClassCode	., 0, 1, 2, 5 or 9	error
		If LengthClassCode = 0 then MinLengthClass must be half cm units	error
		If LengthClassCode = 2 then MinLengthClass must be 2 cm units	error
60	MinLengthClass	If LengthClassCode = 5 then MinLengthClass must be 5 cm units and > 60	error
		If species, LengthClassCode and MinLengthCode are OK then max length is checked	error
		If LengthClassCode = 0 then MinLengthClass must be half cm units	error
		Species referring to the same haul must have the same LengthClassCode	error
63	HLNoAtLengthHour	=>1 and <=999, spaces	error
		=>1 and <=9999999, spaces	error
69	Sex	For each species; NoHour = sum of HLNoAtLengthHour	error
		M or F, spaces	error

HE Haul Information Ocean

3	Quarter	Records referring to the same haul must contain identical values for Quarter	error
4	Country	Records referring to the same haul must contain identical values for Country	error
11	Gear	Records referring to the same haul must contain identical values for Gear	error
21	StationNumber	Records referring to the same haul must contain identical values for StatNo	error
41	Towing Direction	=>0 and <=360, spaces	error
44	GroundSpeed	=>20 and <=60, spaces	error
46	SeedThroughWater	=>10 and <=99, spaces	error
48	WingSpread	=>12 and <=30, spaces	error
50	SurfaceCurrentDirection	=>0 and <=360, spaces	error
53	SurfaceCurrentSpeed	=>0 and <=100, spaces	error
56	BottomCurrentDirection	=>0 and <=360, spaces	error
59	ButtomCurrentSpeed	=>0 and <=100, spaces	error
62	WindDirection	=>0 and <=360, spaces	error
65	WindSpeed	=>0 and <=100, spaces	error
68	SweelDirection	=>0 and <=360, spaces	error
71	SweelHeight	=>0 and <=999, spaces	error

CA SMALK

1	RecordType	If HH has a CA; If year, ship, haul no and species code are the same as previous recorded then quarter, country, gear, rig, doortype and st. no must also be the same	error
		Only on record can be given per haul, species, length class, sex, maturity and age group	error
3	Quarter	Records referring to the same haul must contain identical values for Quarter	error
4	Country	Records referring to the same haul must contain identical values for Country	error
11	Gear	Records referring to the same haul must contain identical values for Gear	error
21	StationNumber	Records referring to the same haul must contain identical values for StatNo	error
42	Area(SubDivRFish)	Check valid list code	error
		SubDiv and StatRec must be consistent and legal	error
44	StatRec	Check valid list code	error
52	LengthClassCode	If SpeciesCode = herring or sprat then LengthClassCode = ., 9 or 0	error
		Species referring to the same haul must have the same LengthClassCode	error
		If LengthClassCode = 5 then MinLengthClass must be 5 cm units and > 60	error
		If LengthClassCode = 2 then MinLengthClass must be 2 cm units	error
		If LengthClassCode = 0 then MinLengthClass must be half cm units	error
		., 0, 1, 2, 5 or 9	error
53	MinLengthClass	If species, LengthClassCode and MinLengthCode are OK then max length is checked	error
56	Sex	M or F, spaces	error
57	Maturity	=>1 and <=4, spaces	error
58	PlusGP	+, spaces	error
59	Age	=>0 and <=99, spaces	error
		If age = space then PlusGr = space	error
		Per haul, species and length class only the oldest age group can be PlusGR	error
61	CANumberAtLengthHour	=>1 and <=999, spaces	error
64	MeanWeight	=>0 and <=999999, spaces	error

APPENDIX IV: CALCULATIONS IN THE BITS DATABASE
(Extract from the document; Description of the BITS database)

3. Calculation of mean CPUE

3.1. Mean CPUE, general

The mean-CPUE is created in two places in the database, in the part where number per age and length (ALK) are generated and in the part where only mean CPUE are created.

Mean CPUE can be calculated for the standard species and the standard by catch species (see the BITS manual, app. VI).

Criteria for calculation of mean CPUE:

- Valid hauls
- Hauls with std. species recording code 1,3 and 4
- Valid species
- No missing length values

Selections:

- Year
- Quarter. If there is selected more than one quarter data will be aggregated into a pseudo quarter. For example, quarter 1 and 2 will be transformed into quarter 12.
- Ship. It is possible to select all ships with in a year or a selection of ships
- Gear. It is possible to select all ships with in a year or a selection of gears
- Sub-division
- Species
- Day/night. It is possible to select day, night or both day and night hauls.
- Fishing power. The user is asked if fishing power should be used in the calculation. This restricts the hauls included to taken by ships with known fishing power. In the present implementation the option only applies to cod only.

3.2. Mean CPUE per length class, calculated per depth stratum

Calculations:

- If the user has chosen to apply fishing power: Number at length per haul is divided with fishing power (*fp*) (Fishing power has only been implemented on vessel surveying in sub-division 25, 26 and 28).

$$mCPUE_{VH,l}^{fp\ corrected} = mCPUE_{VH,l} / fp$$

- Number per length (*l*) (1 cm group) per haul is summed by year, quarter, sub-division and depth stratum
- Number per length is divided with number of valid hauls (*VH*) in the depth stratum

$$mCPUE_l^{fp\ corrected} = \frac{\sum_{VH,l} mCPUE_{VH,l}^{fp\ corrected}}{VH}$$

Created table:

- Mean CPUE per length class and depth stratum (MeanCPUE_ST_Lngt)

3.3. Mean CPUE, calculated per depth stratum

Is calculated from mean CPUE per length class and depth stratum.

Calculations:

- Mean CPUE per length as derived above are summed over length class. This apply to both fp corrected and uncorrected mCpue.

$$mCPUE = \sum_l mCPUE_l$$

Created table:

- Total mean CPUE per depth stratum (MeanCPUE_ST_Total)

3.4. Mean CPUE per length class, calculated per sub-division

Is calculated from mean CPUE per length class and depth stratum.

Criteria:

- Depth stratum '0-19' is deselected

Calculations:

- Mean CPUE per length and stratum is multiplied by the area of the stratum (nm²)
- The multiplied values are summed over stratum
- The summed values are divided with sum of stratum areas

$$mCPUE_l = \frac{\sum_{stratum} mCPUE_{stratum,l} * A_{st}}{\sum_{stratum} A_{st}}$$

Created table:

- Mean CPUE per length class and sub-division (MeanCPUE_SD_Lngt)

3.5. Mean CPUE, calculated per sub-division

Is calculated from mean CPUE per length class and sub-division

Calculations:

- Mean CPUE per length and sub-division are summed over length class

$$mCPUE = \sum_l mCPUE_l$$

Created table:

- Mean CPUE per sub-division (MeanCPUE_SD_Total)

4. Calculation of number per age and cod indices

Number per age and length by either sub-division or depth stratum can be created for the species cod, flounder, herring and sprat. In addition, indices for cod in sub-division 22+24 and 25+25+28 and number per age and length by ship and haul for both cod and flounder can be created.

When extracting the requested ALK from the database the mean CPUE per length class and depth stratum is first calculated as described in the previous section. Afterwards the raw ALK is extracted. The aggregations level is sub-division. Finally the ALK is smoothed and weighted with mean CPUE per length class and depth stratum. The smoothing algorithm is mainly developed for – and tested on cod data. Therefore, the user should careful check the output if the smoothing is selected for any other species. However, it is possible to decide to use the raw ALK when calculating number per age.

Cod indices are only created if the exact sub-divisions are selected and number per age and length by ship is only created if the user requests it.

4.1. Number at age and length (raw ALK)

Criteria for extraction of ALK:

- No missing age values
- Both valid and invalid data

Selections:

- Year
- Quarter. If there is selected more than one quarter data will be aggregated into a pseudo quarter. For example, quarter 1 and 2 will be transformed into quarter 12.
- Ship. It is possible to select all ships with in a year or a selection of ships
- Gear. It is possible to select all gears with in a year or a selection of gears
- Sub-division
- Species
- Day/night. It is possible to select day, night or both day and night hauls.
- Fishing power. The user is asked if fishing power should be used in the calculation. This restricts the used hauls to the ships with known fishing power and is so far for cod only.
- Plus gr. is first used when ALK is fitted. See the following section for more information.

Calculations:

- Number per age (a) and length (l) by haul (H) is summed by year, quarter and sub-division to an age/length key.

$$ALK_{a,l} = \sum_H ALK_{H,a,l}$$

4.2. Fitting of ALK

When ALK is extracted it can be smoothed and weighted by the mean CPUE.

The fitting is done in two steps

- The usual estimator of the age composition is $mCpue_{.a} = mCpue_{.l} * \sum_l \frac{mCpue_l}{mCpue_{.l}} * \frac{ALK_{al}}{ALK_{.l}}$
- In some cases $mCpue_l > 0$ while $ALK_{.l} = 0$ and the estimator cannot be used
- Therefore a new variable $\varphi_{la} = mCpue_l * \frac{ALK_{al}}{ALK_{.l}}$ that is an estimate of the population in length group l of age a in Cpue units
- φ_{la} is continuous in length and therefore it is reasonable to smooth φ_{la} for given a over length. This smoothing is done using a lowess smoother over 5 points.

Tests and value definitions:

- Plus group. The plus group is a sum of all older recorded age groups in the ALK. The minimum plus group is 5 years. If there is data in the database with a plus group less than 5 years the program is aborted.
- Max length. The default and maximal allowed length for the standard species are:
 - Cod: 150 cm
 - Flounder: 65 cm
 - Herring: 600 mm
 - Sprat: 300 mm

If necessary, it is possible to change the maximum length to a length less than default when running the fitting program. The program will then sum all number of fish with length greater than the maximum value into one max length class.
- Quarter and first age group. If the quarter is first quarter or a combination between first and second quarter the first age group is set to one else first age group is zero.

Calculations: Smoothing the length distributions for given age

- The procedure is to make linear regression over five points
 $y = \{\varphi_{l-2,a}, \varphi_{l-1,a}, \varphi_{la}, \varphi_{l+1,a}, \varphi_{l+2,a}\}$, $x = \{l-2, l-1, l, l+1, l+2\}$ giving slope and intercept and using the regression value $\hat{\varphi}_{la} = \text{slope} * l + \text{intercept}$ as the smoothed value. There may be missing observations in the 5 point series and in those cases the point is ignored and the regression is done using only 4 or less points as the case may be. The two first and the two last length groups cannot be smoothed using this procedure and are set to 0.
- The $\hat{\varphi}_{la}$ is normalised so $\sum_{la} \hat{\varphi}_{la} = \sum_{la} mCpue_{la}$
- If the length is less than 15 cm (cod) and there is no ALK the length group is assigned the lowest age, i.e. age 1 for first quarter survey and age 0 for 4th quarter survey.

Created tables:

- Number per hour per age and length per depth stratum (Table: tblSmoothedALK_ST_Lngt)
- Number per hour per age per depth stratum (Table: tblSmoothedALKsum)
- Number per hour per age and length per sub-division (Table: tblSmoothedALK_SD)
- Number per hour per age per sub-division (Table: tblSmoothedALK_SDsum)

4.3. Index calculation

Calculations:

$$i_a = \sum_l \hat{\varphi}_{la}$$

Created tables:

- If sub-division is 22 + 24: tblIndexOne
- If sub-division is 25 + 26 + 28: tblIndexTwo

4.4. Number per age and length by haul

CPUE per age (a) and length (l) by haul (H) are calculated by multiplying number per length in the haul with the fitted ALK ($m_{a,l}$) and divide it with the total number at length in the age group in the fitted ALK ($\hat{\phi}_{l,a}$).

$$CPUE_{H,a,l} = \frac{CPUE_{H,l} * \hat{\phi}_{a,l}}{\hat{\phi}_l}$$

Created table:

- Catch per haul per age group, CHA-file

APPENDIX V: CALCULATIONS IN THE IBTS DATABASE
(Based on the document: IYFS standard tables and matrices)

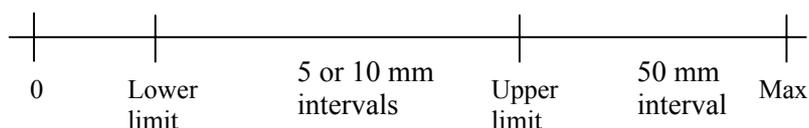
Calculations in the IBTS database are only referring to the standard species.

Basic concepts and variables

Variable	Significance
a	Index for sampling area
r	Index for statistical rectangle
i	Index for length class
j	Index for age group
H(r)	Number of valid hauls in statistical rectangle [r]
C(r)	Number per hour per haul in statistical rectangle [r]
C(r,i)	Length distribution in rectangle [r]: Number per hour per haul in length class [i] and statistical rectangle [r]
C(r,j)	Age distribution in rectangle [r]: Number per hour per haul in age group [j] and statistical rectangle [r]
C(r,i,j)	Age/length distribution in rectangle [r]: Number per hour per haul in length class [i], age group [j] and statistical rectangle [r]
f(a,i,j)	Age/length key for length class [i] in sampling area [a]: The proportions of fish in length class [i] falling in age group [j]
O(a,i,j)	Number of otoliths sampled in sampling area [a], length class [i] and age group [j]
a	Index for sampling area
H(a)	Number of valid hauls in sampling area [a]
C(a)	Number per hour per haul in sampling area [a] (total of all rectangles in the area)
R(a)	Number of rectangles sampled in sampling area [a]
Cr(a)	Mean number per hour per haul in sampling area [a] (mean of rectangles in the area)
C(a,i)	Length distribution in sampling area [a]: Number per hour per haul in length class [i] and sampling area [a] (total of all rectangles in the area)
C(a,j)	Age distribution in sampling area [a]: Number per hour per haul in age group [j] and sampling area [a] (total of all rectangles in the area)
Cr(a,j)	Mean distribution in sampling area [a]: Mean number per hour per haul in age group [j] and sampling area [a] (mean of all rectangles in the area)
C(a,i,j)	Age/length distribution in sampling area [a]: Number per hour per haul in length class [i], age group [j] and sampling area [a] (total of all rectangles in the area)

Calculating mean CPUE per length class

In the IBTS database the basic unit is mean number per hour per length class in a statistical rectangle. The mean value is only calculated from valid hauls and a stat. rec. is excluded from further analysis if number of hauls in the stat. rec. is zero. The CPUE value per length class is adjusted with values from hauls where only total number per haul is measured. The recorded length classes are transformed into a new variable having a grouping for low and high values:



The limits and interval size depend on species. A fish will be included in a plus group if it is greater than the maximum length (table 1).

Table 1. Lower and upper limits used when creating the length distribution.

Rubin code	Lower limit (mm)	Upper limit (mm)
POLL VIR		600
CLUP HAR		400
SCOM SCO		600
SPRA SPR		200
GADU MOR		600
MELA AEG		600
MERL MNG		600
TRIS ESM		300

Adjustment of catches data to a common basis

In the IBTS database the following haul and catch variables are defined:

Variable	Significance
NoHauls	Number of hauls in district, exclusive of HaulsNHr
HaulsNHr	Number of hauls where only number per hour was measured
InvExcl	Number of invalid hauls – species excluded
NghtExcl	Number of night hauls excluded
NoInDist	Number of fish in district, exclusive of FishNHr
FishNHr	Number of fish where only number per hour was measured
NoAtLngt	Number caught per hour for a length class

Number of $H(r)$ in statistical rectangle $[r]$ comprises the valid hauls:

$$(1) \quad H(r) = \text{NoHauls} + \text{HaulsNHr}$$

For all catch data the following adjustment factor is computed:

$$(2) \quad \text{AdjFac} = \frac{\text{NoInDist} + \text{FishNHr}}{\text{NoInDist}} * \frac{1}{\text{NoHauls} + \text{HaulsNHr}}$$

and the Number per hour per haul $C(r)$ is then computed as:

$$(3) \quad C(r) = \text{AdjFac} * \text{NoInDist}$$

So, AdjFac performs the following adjustments:

- Relative upgrading for FishNHr, i.e. the hauls where only number per hour was measured.
- Division by the number of hauls $H(r)$, so that $C(r)$ represents an average of all the hauls in the statistical rectangle.

The AdjFac is also used for adjusting the length distribution data where, by its very nature, the catch represented by FishNHR is not included.

The length distribution $C(r,i)$ is obtained from NoAtLngt by

$$C(r,i) = AdjFac * NoAtLngt$$

By using the same adjustment factor, it is assured that $C(r)$ and $C(r,i)$ refer to the same basis, that is the average (by hauls) number of fish caught per hour per haul in the statistical rectangle.

Age/length keys

Otolith keys

The normal otolith age/length keys is a straightforward summation and normalisation of the distribution:

$$f(a,i,j) = \frac{O(a,i,j)}{\sum_j O(a,i,j)}$$

The standard age group interval is from 1 to 6 years. Hence, the upper limit of the age distribution is usually 6 years.

Substitution areas

Substitution areas are sampling areas used as substitutions for other areas, where the number and/or quality of the sampled otoliths is insufficient to provide the basis for usable age/length keys.

The substitution areas are input manually into the database, based on the following criteria:

1. If the number of otoliths in a sampling area is less than 25, the age/length key is considered insufficient, so that it should be supplemented or replaced by keys from substitution areas having a greater number of otoliths.
2. For roundfish area 8 and 9 no substitution areas are used, i.e. age distributions are not computed if proper otolith keys are not available.
3. If none of these criteria can be applied, either no substitution areas are used (i.e. no age distributions), or the closer areas with proper keys are used as distribution areas. In case of very few otoliths sampled all together (e.g. for mackerel), a key common for the whole area may be formed by pooling all otoliths sampled.

Age distribution

The age distribution $C(r,j)$ is computed by dividing the length distribution $C(r,i)$ according to the age/length keys $f(a,i,j)$:

$$C(r,i,j) = C(r,i) * f(a,i,j)$$

$$C(r,j) = \sum_i C(r,i,j)$$

The computational problem in calculating $C(r,j)$ is what to do if

- there are no age/length keys available in the area, or
- the available age/length keys do not match with the catch in a given length class

The first problem is solved as explained earlier by using substitution areas or, if that is not possible, excluding all rectangles for which there are no age/length keys available in the database.

Concerning the second problem, we can assume that there is at least one age/length key available for the statistical rectangle concerned (otherwise, the first category applies), and the computational situation can be described as follows:

1. If a key for the actual length class [i] exist this key is used.
2. If it is not possible to find a key for exactly the actual length class [i], the program takes the length class closest to [i]. If two keys have the same distance to [i] the average of the keys is taken.
3. As described previously there are at least 25 otoliths behind the age/length key in a given sampling area. In most cases this should assure a reasonable coverage of all length classes [i]. At either end of the length class scale, i.e. very small respectively very large length classes, the above procedure may give gross erroneous results if the available keys do not cover these extremes. The result is that very young and/or very old fish are grouped incorrectly.
4. To reduce the number of instances, where either of these situations lead to flawed age distributions, the program uses some species dependant threshold values for the length of the lowest and highest age group (table 2).

Table 2. Species dependant values for the length of the lowest and highest age group (in mm).

Rubin code	Minimum acceptable age as plus group	Length split for 1-group, 1. quarter (l_1)	Length split for 0-group, 2. quarter (l_1)	Length split for plus group, 1. quarter (l_2)	Length split for plus group, 2. quarter (l_2)	Length split for plus group, 2. quarter (l_2)	Length split for plus group, 4. quarter (l_2)
POLL VIR	6	250	70	900	1100	1100	1100
CLUP HAR	5	200	60	320	340	340	340
SCOM SCO	6	250	50	450	450	450	450
SPRA SPR	5	100	40	160	180	180	180
GADU MOR	6	250	70	900	1100	1100	1100
MELA AEG	6	200	100	600	700	700	700
MERL MNG	6	200	80	450	500	500	500
TRIS ESM	6	150	50	250	250	250	250

5. The threshold value l_1 is only used if there are no keys less than l_1 : in these cases length classes [i] less than l_1 is put into the lowest age group (i.e. 1). Similarly, the threshold value l_2 is only used if there are no keys greater than l_2 : in these cases length classes [i] greater than l_2 is put into the highest age group (i.e. 5 or 6).

Aggregation of data

Aggregation by area takes place on two levels:

- Sampling area are groupings of stat. rec. into wider areas. At present only roundfish area is allowed.
- The standard area for a particular species is a selection of the statistical rectangle used for computation of abundance index for the species in question. It species depended.

The following formulas describe how to aggregate data on sampling area level. The variables and relationships for aggregation by standard area are entirely parallel to aggregation by sampling area and will therefore not be described.

Number of valid hauls in sampling area:

$$H(a) = \sum_{r \in a} H(r)$$

Number per hour per haul in sampling area:

$$C(a) = \sum_{r \in a} C(r)$$

Number of rectangles sampled in sampling area:

$$R(a) = \sum_{r \in a} 1$$

Mean number per hour per haul in sampling area:

$$Cr(a) = \frac{C(a)}{R(a)}$$

Length distribution in sampling area:

$$C(a, i) = \sum_{r \in a} C(r, i)$$

Age distribution in sampling area:

$$C(a, j) = \sum_{r \in a} C(r, j)$$

Mean distribution in sampling area:

$$Cr(a, j) = \frac{C(a, j)}{R(a)}$$

Age/length distribution in sampling area:

$$C(a, i, j) = \sum_{r \in a} C(r, i, j)$$